

# 特許文書の自動翻訳技術

—特許文書のための高精度の自動翻訳技術を研究開発して、  
日本製品の特許侵害・訴訟リスクの回避を支援します—



隅田 英一郎 (すみた えいいちろう)

ユニバーサルコミュニケーション研究所 多言語翻訳研究室 室長

1982年、大学院修士課程修了。機械翻訳、eラーニングを研究。博士(工学)。

## 1. はじめに

第16代アメリカ合衆国大統領リンカーンの言葉で「特許制度は、天才の火に利益という油を注いだ」が残っています。

発明は、特許制度がなければ、他人に簡単に盗まれてしまいます。特許制度は、こういったことが起こらないよう、発明者には一定期間、独占的な権利を与えて保護を図るものです。製品を作り販売するためには、他人の特許を侵害しないように、予め調査しておく必要があります。これを怠って、侵害してしまうと、裁判になり、膨大な補償金を払うこととなります。

特許制度は国ごとに定められており、日本では日本語、韓国では韓国語、中国では中国語で各国政府に申請することになっています。一方、経済はグローバル化していますから、例えば、日本の企業が中国に製品を輸出するためには、中国の特許の調査が必要になります。中国は今や世界第2位の経済大国ですので、日本企業も中国市場への進出が今後の発展の要になります。一方で、最近の中国ではその特許の出願数も急速に伸びており、今や世界第1位です。実際に侵害・訴訟事案が増加しています<sup>[1]</sup>。困ったことに、中国語を日本語に翻訳ができる翻訳者の数は限られますし、人間による翻訳はコストと時間がかさみます。そこで、中国語特許文書の高精度の自動翻訳システムの開発が焦眉の課題となっていました。

## 2. 対訳コーパスに基づく自動翻訳技術

ハードウェアの処理速度や記憶容量が格段に進歩したことなどから、文章や辞書が大量に計算機上に集積されるようになったことなどから、自動翻訳の研究において、対訳コーパス(同じ意味の原文と訳文の文レベルの対を集めたもの)から、翻訳に必要な知識を自動的に構築する技術が興り、現在、主流の研究手法となっています。NICTは既に、旅行会話やeコマースの分野で多言語翻訳システムを実現し、高い翻訳品質

を達成しています<sup>[2]</sup>。

この方法では、①ある一定量以上の対訳コーパスが集まると翻訳品質が実用レベルになること、量が多ければ多いほど品質が高くなるがわかっていますので、対訳コーパスを経済的に短期間で収集する手法を確立することが重要になります。また、②対訳コーパスが同じデータ量でもアルゴリズムによる性能差が大きいことがわかっていますので、与えられたデータで高精度を実現する良いアルゴリズムの研究が重要になります。

## 3. 特許翻訳に内在する重要な研究課題

特許文を翻訳することは大変難しいです。実際、翻訳会社の翻訳費用の単価も他の分野の文書より大幅に高額になっています。

理由の1つ目は一文の長さが非常に長いことによって、解釈が困難になり翻訳誤りが増えること、2つ目は専門用語が膨大で、これを十分カバーする対訳辞書が存在しないこと、関連して、翻訳するために内容理解が必要でそのために専門的な背景知識が求められることがあります。さらに、中国語と日本語は文法が全く異なる(細部にこだわらなければ、中国語の文法は英語の文法に近いと考えられ、中国語と日本語とは語順が全く異なります。)ことから、従来中国語から日本語への自動翻訳技術では翻訳精度が低い状況でした(表1に示した従来技術A~Cのように、意味不明な翻訳が出力されることが少なくありません)。

表1●中国語の特許文の例とその翻訳

中国語の原文	图一表示应用本发明的车用发动机的传感器设置结构的发动机一的整体结构的图
従来技術Aの訳	図はちょっと本発明した車を応用することを示してエンジンでのセンサーであり構造のエンジンの1の全体構造の図を設置します
従来技術Bの訳	本発明の最初の車は、図の全体的な構造を提供するエンジンセンサーのエンジンの構造に適用されます
従来技術Cの訳	図は1つは応用の当発明の自動車用エンジンのセンサーが構造のエンジン1の全体の構造の図を設けると表しています
提案技術での訳	図一は本発明に係る車両用エンジンのセンサ配設構造のエンジン一-の全体構成を示す図
模範訳	図1は、本発明に係る車両用エンジンのセンサ配設構造を応用したエンジン1の全体構成図を示している

#### 4. 特許翻訳の高精度自動翻訳を目指して

NICTでは、この困難な研究課題を解くために、**「2.対訳コーパスに基づく自動翻訳技術」**①の研究を進め、特許に関しては大規模な対訳コーパスを構築してきました。さらに、②の研究を進め、「訳語選択に関する翻訳知識」と「語順の変更に関する翻訳知識」を自動的に獲得する新たなアルゴリズムを創出しました。さらに、日本特許情報機構（Japio）との共同研究を行うことで、自動翻訳システムの訳文に対して、特許のプロの目で見えた「通じる翻訳」に向けた評価をしてもらうことができ、効率的な改良を行うことができました。新技術に基づいて開発した、翻訳者が判定した「中日自動翻訳ソフトウェア」の精度は、従来技術の3倍以上の値を達成しています。図1のように、この「中日自動翻訳ソフトウェア」によって、Japioは中国の特許文献を日本語に翻訳及びデータベース化し、来春にも、有償サービス「Japio世界特許情報検索サービス（Japio-Global Patent Gateway）」<sup>[3]</sup>の拡張版として事業化する計画です。企業の知財部や弁理士の知財調査や特許庁審査官の先行技術調査に役立って、日本の企業の特許侵害のリスクを軽減できると期待します。

#### 5. おわりに

NICTは、今後も、翻訳精度と翻訳速度の一層の向上のために研究を継続していきます。

また、一文が長く専門用語が多く、非常に翻訳が困難な特許での成功は、他の分野での中日自動翻訳の研究開発を加速すると考えられます。

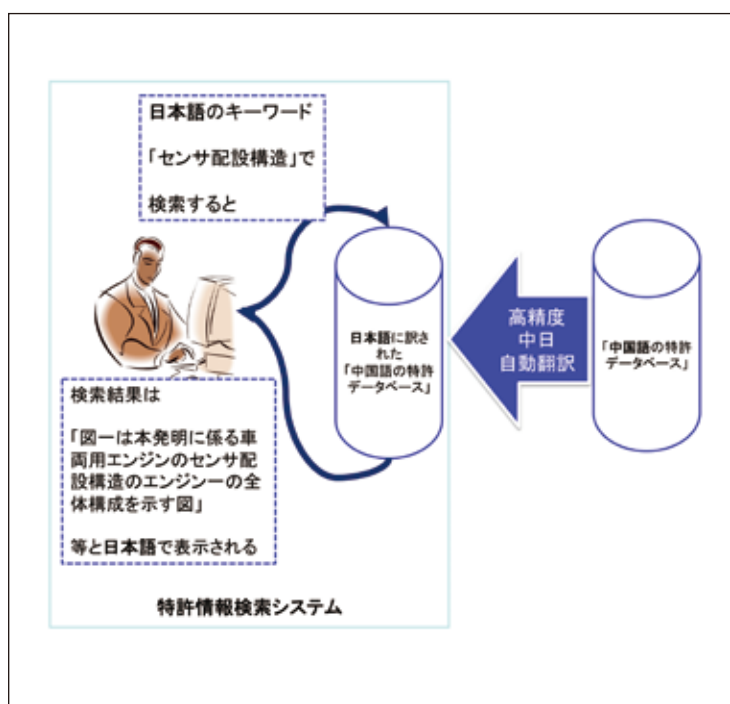


図1●自動翻訳を活用した中国語特許の検索システム

#### 参考資料

- [1] 産業構造審議会 知的財産政策部会 平成24年6月25日配布資料「知財立国に向けた新たな課題と対応」  
[http://www.jpo.go.jp/cgi/link.cgi?url=/shiryou/toushin/shingikai/sangyou\\_kouzou.htm](http://www.jpo.go.jp/cgi/link.cgi?url=/shiryou/toushin/shingikai/sangyou_kouzou.htm)
- [2] 隅田 英一郎「多言語自動翻訳技術」NICT NEWS 2011年3月号  
<http://www.nict.go.jp/publication/NICT-News/1103/02.html>
- [3] Japio世界特許情報検索サービス（Japio-GPG）  
<http://www.japio.or.jp/service/service05.html>