

中国語特許翻訳を支援する 機械翻訳技術

東芝ソリューション株式会社 プラットフォームソリューション事業部参事 **熊野 明**

PROFILE

1982年東京工業大学卒業。同年東京芝浦電気(株)(現(株)東芝)入社。2010年から東芝ソリューション(株)プラットフォームソリューション事業部参事。自然言語処理システムの研究開発に従事。アジア太平洋機械翻訳協会理事。AAMT/Japio 特許翻訳研究会委員。2007年度から特許版・産業日本語委員会委員。

✉ kumano.akira@toshiba-sol.co.jp

1 はじめに

近年、中国での技術開発、製品販売、ビジネス展開が急速に進むにつれ、中国への特許出願が急増している。2010年には、日本国内での特許出願数を上回った。日本企業からの中国特許出願数も、大きく伸びている。

我々日本人が中国への特許出願を行う際、国内への特許出願に比べて大きな障害がある。

その一つは、中国の特許調査における情報検索である。一定のキーワードや分野情報で絞り込むことはできるが、検索された特許を正確に理解することは容易ではない。中国語の特許の日本語への翻訳を依頼することもできるが、コストが課題である。

もう一つは、中国語での特許出願である。日本国内の技術者で、中国語で特許明細書を執筆できる人は少ないであろう。ただ、中国に出願する特許の多くは、その前に日本国内で出願したもの、あるいは同時に日本国内出願するものである。したがって、日本語の明細書は作成済みである。この日本語明細書に対して、中国語への翻訳を依頼するのが近道であるが、やはりコストが課題である。

我々は、従来、英日・日英機械翻訳において、特許文書用の機能を開発し、提供してきた[1]。さらに、上で述べた中国語特許の課題を軽減できるように、中日・日中機械翻訳を開発した。本稿では、中国語特許文書の翻訳に関わる技術について紹介する。これらの機能は、The 翻訳シリーズの中日・日中機械翻訳製品[2][3]で

提供している。

2 中国特許調査のための 機械翻訳

中国語の特許を検索、理解するために、中日機械翻訳を利用することは可能であるが、一般的には多くの困難がある。新語や専門用語を多く含む特許文を翻訳する場合、専門用語辞書の拡充は不可欠である。しかし、辞書の拡充だけでは解決できない問題も多い。

2.1 中国語特許文の解析

現在の機械翻訳に求められるものは、原文の意味を正しく訳文に表現することである。訳文の表現が多少不自然でも、原文の内容が誤りなく出力されていることが望ましい。つまり、規則ベースの機械翻訳では、原文解析の精度が翻訳品質を大きく左右する。

中日機械翻訳の原文解析における大きな課題は、単語分割と品詞付与である[4]。中国語の文はすべて漢字で表記され、単語の間に空白は存在しない。したがって、単語分割の曖昧性が常に問題となる。さらに、同じ表記の単語でも複数の品詞を持つものがある。名詞の多くは動詞としても機能する。動詞の一部には、機能語や副詞にもなるものもある。正しい品詞を付与するために、統計的情報を利用するが、原文1文の情報だけでは、正しく解決できない場合もある。

2.2 文書構造を利用した中日翻訳

特許明細書に記述した発明の内容を的確に表現したものが、発明の名称である。技術の高度化、細分化に伴い、発明の名称は、構成要素や特徴を記述した長い表現になる傾向がある。長い表現になれば、構文解析の曖昧性は増大する。一定以上の語数を含む表現では、全体が名詞句として解釈できるだけでなく、動詞句として解釈できる場合が多い。さらに介詞句として解釈できる表現もある。

図1は、中国特許の発明の名称と発明者の記述例である。

発明名称	基于加密算法技术的商品防伪方法
発明人	龙传红

図1 中国特許の記述例（一部）

発明の名称や発明者の記述であるという文書構造の知識を利用しないで、通常の文として翻訳した結果を図2に示す。

発明の名称	暗号化アルゴリズム技術の商品偽造防止方法に基づいて
発明者	龍は伝わって赤くなる

図2 中日翻訳結果の例

発明の名称は名詞句として解釈すべきだが、介詞句として訳されている。単純な中国語表現の日本語訳としては可能性があるが、発明の名称の翻訳結果としては正しくない。

このような不具合を軽減するため、文書構造を利用して翻訳を行うことができる。

上述の中国特許の発明の名称や発明者の記述は、HTMLやXMLのタグによって文書構造が明確に示されている。図1で示した部分のHTML記述例を、図3に示す。

```
<TR>
  <TD> 発明名称 </TD>
  <TD> 基于加密算法技术的商品防伪方法 </TD>
</TR>
<TR>
  <TD> 発明人 </TD>
  <TD> 龙传红 </TD>
</TR>
```

図3. 中国特許のHTML記述例（一部）

このような構造表現を利用することで、「基于加密算法技术的商品防伪方法」は「発明名称」（＝発明の名称）に対応する部分であることが分かる。この表現に対しては、文全体を名詞句としての解釈を優先して翻訳することにより、大きな誤訳を避けることができる。

同様に、発明者の記述部分も、構造を把握することができる。「龙传红」が「発明人」（＝発明者）に対応することが分かるので、人名を収録した固有名詞辞書の見出し語を優先することにより、人名としての解釈を優先して翻訳することができる。

上述した文書構造を利用して中日翻訳した結果を図4に示す。発明の名称は名詞句として、発明者は人名として翻訳できる。

発明の名称	暗号化アルゴリズム技術に基づく商品偽造防止方法
発明者	竜伝紅

図4. 文書構造を利用した中日翻訳結果

3 中国語明細書作成のための機械翻訳

中国に特許出願する多くの場合に共通する業務は、日本語特許明細書の中国語への翻訳である。機械翻訳を利用することで、高速かつ均質に翻訳することができる。しかし、長い原文に対しては、一般に構文解析の曖昧性が大きくなり、曖昧性を正しく解消できないと、十分な

翻訳精度が得られない。これは日中翻訳だけの問題ではなく、日英翻訳にも共通する問題である。

3.1 日本語特許長文の解析

日本語の特許明細書に現れやすい長文表現の一例を図5に示す。ある装置を構成する4つの要素を、並列表現で示したものである。

【構成】

重量検出装置1において、ターンテーブル18と、このターンテーブル18上の物体19の重量を荷重として受け、この荷重を負荷トルクに変換しながらターンテーブル18を回転自在に支持する支持手段20と、前記ターンテーブル18を負荷トルクに対応した駆動トルクで回転駆動するモータ3と、このモータ3の負荷トルクを検出して負荷トルクと荷重との相関関係からターンテーブル18上の物体19の重量を間接的に測定する荷重測定装置2とからなる構成とした。

図5. 日本語特許の記述例

このまま日本語を構文解析して翻訳すると、1文としての解釈に失敗し、自動的に解析可能な部分に分割する。その結果、図6に示すように、原文の並列構造を反映しない断片的な訳文を出力してしまう。

【構成】

放重量查出装置1, 作为负荷接受转盘18和这个转盘18上的物体19的重量, 把这个负荷一边转换为负荷转动力矩一边转自由自在的支持转盘18的支撑手段20,

由在对应负荷转动力矩的驱动转动力矩旋转驱动上述转盘18的电动机3和查出这台电动机3的负荷转动力矩从负荷转动力矩和负荷的相互关系间接地测量转盘18上的物体18的重量的负荷测量装置2变成的构成和下面。

図6. 原文に対する日中翻訳結果

3.2 長文の自動前編集による分割

構成要素を列挙した長文を正しく翻訳するために、構成要素ごとに自動分割する。分割には、日本語明細書の表層的特徴を利用した前編集規則を適用する。この例で適用した規則を、図7に示す。〈NPn〉は、任意の名詞句を表す。

[原文パターン (入力)]

〈NP1〉において、〈NP2〉と、〈NP3〉と、〈NP4〉と、〈NP5〉とからなる構成とした。

↓

[前編集パターン (出力)]

〈NP1〉に、以下を備えて構成する。

〈NP2〉

〈NP3〉

〈NP4〉

〈NP5〉

図7. 長文分割用の前編集規則の例

この手法は、日英機械翻訳用の自動前編集と同じである[5]。自動前編集による分割結果を図8に示す。

【構成】

重量検出装置1に、以下を備えて構成する。

①ターンテーブル18

②このターンテーブル18上の物体19の重量を荷重として受け、この荷重を負荷トルクに変換しながらターンテーブル18を回転自在に支持する支持手段20

③前記ターンテーブル18を負荷トルクに対応した駆動トルクで回転駆動するモータ3

④このモータ3の負荷トルクを検出して負荷トルクと荷重との相関関係からターンテーブル18上の物体19の重量を間接的に測定する荷重測定装置2

図8. 日本語特許の前編集結果

各行頭の①～④は、前編集結果の出力に含まれる文字ではなく、構成要素を分かりやすくするためのものである。

自動前編集による分割結果に対して日中翻訳したものを図9に示す。構成要素ごとに翻訳された結果、原文の意味を正しく伝える中国語表現を出力することができる。

【構成】

重量查出装置 1, 具有以下構成：

- ①转盘 18,
- ②作为负荷接受这个转盘 18 上的物 19 的重量, 把这个负荷一边转换为负荷转动动力矩一边转自由自在的支持转盘 18 的支持手段 20
- ③在对应负荷转动动力矩了的驱动转动动力矩旋转驱动上述转盘 18 的电动机 3
- ④查出这台电动机 3 的负荷转动动力矩从负荷转动动力矩和负荷的相互关系间接地测量转盘 18 上的物 19 的重量的负荷测量装置 2

図9. 前編集結果に対する日中翻訳結果

参考文献

- [1] 熊野：知的財産のグローバル化を加速する機械翻訳技術，東芝レビュー Vol.64 No.2, pp.10-13, (2009)
- [2] 東芝ソリューション(株): The 翻訳エンタープライズ
<http://mt-server.toshiba-sol.co.jp/> (2011)
- [3] 東芝ソリューション(株): 情報活用 SaaS 翻訳「Eiplaza/MT」
<http://www.toshiba-sol.co.jp/sol/eiplaza/jouhou03.htm> (2011)
- [4] 出羽，熊野：中日・日中機械翻訳システム，東芝レビュー Vol.62 No.4, pp.30-33, (2007)
- [5] 鈴木，熊野：特許文書用前編集機能を備えた機械翻訳システム，情報処理学会第63回全国大会II, pp.255-256, (2001)

4 まとめ

中国語特許の調査や出願を支援する機械翻訳技術について述べた。本機能は、The 翻訳エンタープライズ、SaaS Eiplaza/MT の中日・日中機械翻訳エンジンに実装している。

中国語特許の翻訳を支援するための機械翻訳には、専門用語辞書の拡充などによる翻訳エンジンの精度向上と、それを活用した特許文書専用機能が必要である。今後も、中国語の特許翻訳に関わる業務をさらに効率化するため、翻訳精度向上、機能開発を行っていく。