

# 産業日本語プロジェクトと 特許版産業日本語の開発

機械翻訳を活用した外国出願明細書の作成コスト削減に向けて

一般財団法人日本特許情報機構  
特許情報研究所調査研究部長

渡邊 豊英

**PROFILE**

1989年 特許庁入庁、2002年 特許庁情報技術企画室調査班長、  
2004年 特許庁審判部審判官、2007年 特許審査第二部繊維包装機械主任上席審査官  
2008年10月より現職。



## 1 はじめに

経済のグローバル化が進展する中で、様々な情報の発信や伝達は、事実上の世界標準言語といえる英語をベースとして行われている。日本語という国際的にはマイナーといわざるを得ない言語を母国語に持つ我が国は、情報の発信や収集を翻訳という負荷を負いながら行うことになる。また、特許の世界では、海外への特許出願や海外特許情報の活用が以前にも増して重要となってきた。これに伴い、海外への特許出願や海外の特許情報を活用するために必要な翻訳コストも膨大なものとなってきている。このような状況の下、一定の翻訳品質を担保しつつ、翻訳コストを少しでも下げることが急務となっており、機械翻訳を有効に活用することの必要性・重要性があらためて認識されつつある。しかしながら、現状の機械翻訳の活用を考える際には、機械翻訳システム側からのアプローチのみでは限界があり、機械翻訳の対象となる原文の書き方において、何らかの取り組みが必要である。特に、特許分野における活用を考える際には、現状の特許関連文書が有する問題点、すなわち、一文が長く、係り受けが複雑で人にとってでさえ分かりづらいこと、用語の統制ができておらず、造語（複合語）が安易に用いられていること等の問題点を無視することはできない。

このような状況へ対応するべく、Japioは、産業日本語プロジェクトの推進に取り組んでいる。産業日本語とは、産業活動の諸側面を媒介する産業技術文書を人に

理解しやすくかつ機械にも処理しやすく記述し表現するための日本語であり、特許版産業日本語とは、特に特許関連文書の記述に適した産業日本語である。そして、産業日本語プロジェクトとは、産業日本語の我が国産業界への導入を支援する仕組みである産業日本語プラットフォームの開発や普及シナリオを策定するプロジェクトである。

本稿では、産業日本語プロジェクトの要である産業日本語プラットフォームの概要と、特許版産業日本語の開発に関する今後の取り組みの概要について紹介する。

## 2 産業日本語プラットフォームの概要

産業日本語プラットフォームは、日本語を情報伝達のベースとする我が国の産業界に、産業日本語という新たな考え方に基づく日本語の枠組みを普及し根付かせるための基盤となる仕組みである。産業日本語プラットフォームを構成する主要な要素について説明する。

### ① 産業日本語共通基盤仕様

産業技術文書には、特許明細書等の特許関連文書をはじめとし、学術論文や設計書、仕様書、マニュアルといった様々なカテゴリの文書が存在する。これらの文書には、技術情報を読み手に対して正確に伝達するという共通の基本的な機能の他に、カテゴリに応じた機能を付加的に具備することが要求される。例えば、特許関連文書の場合には、特許法等の法規に規定される記載要件を満たすという機能が付加的に要求される。

産業日本語を、こうした様々なカテゴリの産業技術文書に対して導入する際には、技術情報を読み手に対して正確に伝達するという共通の基本的な機能を担保し、かつ、カテゴリに応じた詳細仕様を生成するベースとなるべきものが必要である。それが、産業日本語共通基盤仕様である。産業日本語共通基盤仕様は、産業技術文書における記載の明晰性を体系的かつ統一的に定義するための共通の枠組みと、仕様を記述するための共通形式とで構成される。

② オーサリングシステムと言語知識集合知サーバ

産業日本語を我が国産業界に迅速かつ広範に普及させるためには、文書が作成される現場に、産業日本語の仕様に則った文書の作成を支援する仕組みを提供する必要がある。オーサリングシステムと言語知識集合知サーバは、これを実現するためのシステムである。

オーサリングシステムは、産業日本語の仕様に則った明晰な日本語文書の作成を支援するシステムである。ある文書が非明晰な場合、すなわち、何を伝えようとしているのかが不明瞭である場合、その文書を真に明晰化できるのは、文書の作成者自身である。第三者がこの明晰化の作業を行う場合には、その文書の作成意図を推察し、その結果、依然として不明瞭であるときは、その意図を文書作成者に確認する工程が必要である。オーサリングシステムは、この工程を代行するものであり、文書作成者との間でインタラクティブなやりとりをすることにより、入力された文が、意図したとおりの情報を表現しているかを文書作成者に明示し、必要に応じて修正を促すことで、その文の明晰性を担保する。以下に、オーサリングシステムの利用イメージを示す。

まず、オーサリングシステムに、次の文を入力したとする。

以下の説明では、ソース言語の文及び単語をそれぞれソース文及びソース単語、ターゲット言語の文及び単語をそれぞれターゲット文及びターゲット単語と呼ぶ。



オーサリングシステムは、入力文を解析し、最上位文節と係受け関係を以下のように表示しノ格が二重になっていることを警告する。

以下の説明では、ソース言語の文及び単語をそれぞれソース文及びソース単語、ターゲット言語の文及び単語をそれぞれターゲット文及びターゲット単語と呼ぶ。  
<□では、□を□を□と呼ぶ>



ユーザは、文が非明晰であることからシステムが解析誤りをしたと判断し、省略されている「と呼び」を書き加える。

以下の説明では、ソース言語の文及び単語をそれぞれソース文及びソース単語と呼び、ターゲット言語の文及び単語をそれぞれターゲット文及びターゲット単語と呼ぶ。



システムは、修正された文を解析し、以下のような解析結果を表示する。

以下の説明では、ソース言語の文及び単語をそれぞれソース文及びソース単語と呼び、ターゲット言語の文及び単語をそれぞれターゲット文及びターゲット単語と呼ぶ。  
<□では、□を□と呼び、□を□と呼ぶ>



ユーザは、表示された解析結果から、システムが正しく文を解析したと判断し、文を確定する。

以下は、確定された文を、汎用の機械翻訳ソフトで翻訳した例である。



In the following explanation, the sentence and word of a source language are called a source sentence and a source word, respectively, and the sentence and word of a target language are called a target sentence and a target word, respectively.

このように、オーサリングシステムは、インタラクティブな操作を介して、入力されたテキストを、人にとっても明晰であり、機械翻訳等の機械処理にも適した産業日本語テキストに変換する。

オーサリングシステムは、作成対象の文書に対応した産業日本語の仕様に基づく規則や辞書類を入れ換えることによって、様々な産業日本語仕様に対応することができる。

言語知識集合知サーバは、規則や辞書類の内容を、このシステムを運用する分野に関わる人々の集合知によって漸進的に整備していくシステムである。言語知識データベースは、用語や表現等に関する共通化すべき言語知識を共有する仕組みであって、共有単位毎に作成される。言語知識集合知サーバでは、産業日本語に必要な

言語知識データが各言語知識データベースに格納され、データベース作成支援機能とデータベース更新管理機能により、データベースの構築および管理が行われる。また、オーサリングシステムにおける規則・辞書ベース管理機能部と通信を行い、産業日本語に必要な言語知識を収集し、データベースを更新するとともに、オーサリングシステム内の規則・辞書ベースの更新を行う。

### 3

## 特許版産業日本語の開発と今後の取り組み

～機械翻訳を活用した外国出願明細書の作成～

産業日本語プラットフォームの開発のポイントは、確実な産業日本語共通基盤仕様を策定することにある。Japioは、自らの専門性を活かすことができ、また、特許公報として利用可能な膨大な量の言語資源の蓄積がある特許分野から、産業日本語を導入するための研究開発を行い、その効果を実証しつつ、徐々にその適用範囲を拡げていくことを考えている。そのために、Japioは、知財分野の専門家の方々や、自然言語処理分野の専門家

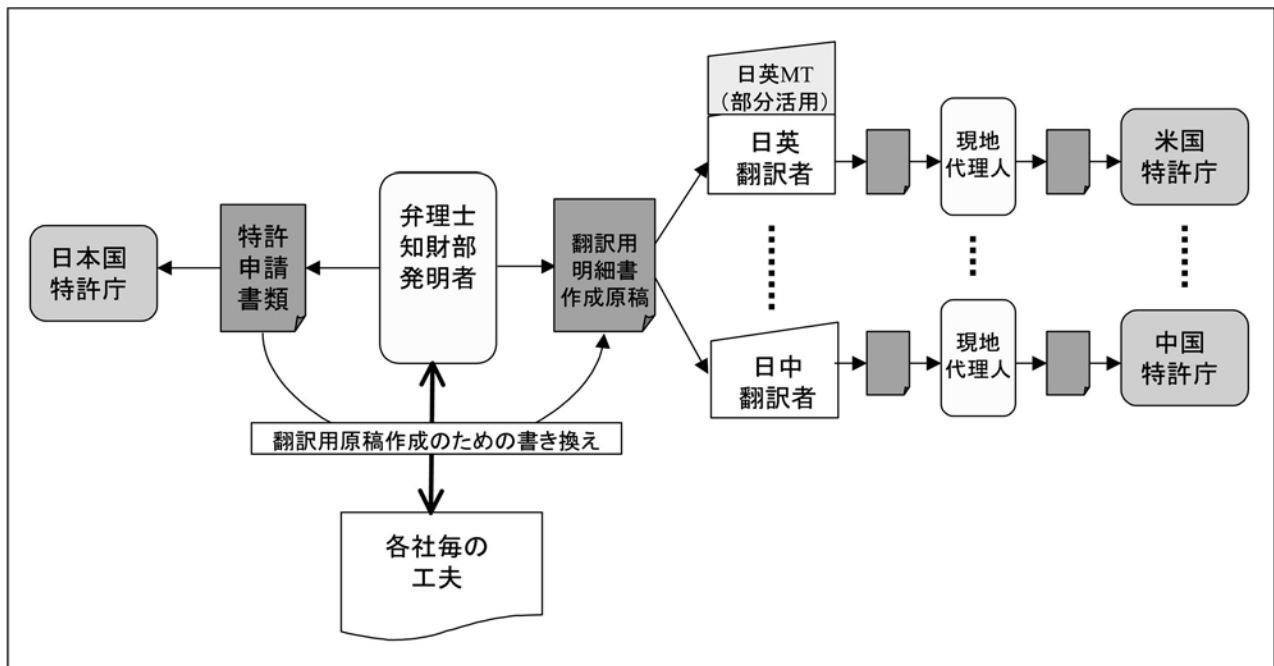


図1 外国出願明細書作成の現状

の方々のご協力を得て特許版産業日本語委員会を設置し、特許版産業日本語の検討と、産業日本語共通基盤仕様の策定作業を行ってきた。平成20年度には、特許版産業日本語の基本仕様（第0版）と、産業日本語共通基盤仕様の第1版を策定した。今後は、特許版産業日本語の詳細仕様を検討しつつ、実用化に向けて、共通基盤仕様第1版を更に改良・改訂していく必要がある。

平成21年度の特許版産業日本語委員会の目標は、特許明細書ライティングマニュアルの作成と、産業日本語共通基盤や特許版産業日本語仕様のブラッシュアップである。

特許明細書ライティングマニュアルの作成における当面の目標は、外国出願明細書の作成時に役立つツールとなることである。現在、多くの企業では、特許庁への出願明細書を、外国出願明細書の下地となる翻訳用原稿の形に書き換え、これを翻訳者により翻訳し、出願先国の代理人等の手により、出願用明細書を作成している。

(図1を参照。)この翻訳用原稿への書き換えの現場に、特許明細書ライティングマニュアルを導入し、書き換えを特許版産業日本語で、機械翻訳が活用できるように行い、かつ、その工程を定式化することで、大幅な効率の向上が期待される。

特許明細書ライティングマニュアルの作成にあたっては、まず、特許明細書が満たすべき記載要件や明細書に記載されるべき情報を整理するとともに、特許明細書の記載に用いられるべきメタ文（メタ表現）を抽出したいと考えている。メタ文の抽出ができれば、それを機械翻訳システムにパターン登録することにより、翻訳精度の向上が見込まれる。また、実事例をもとに、現行の機械翻訳で、意味内容を変えることなく正確に翻訳される文と、そうでない文とを体系的に仕分け、そうでない文を、機械翻訳により、意味内容を変えることなく正確に翻訳されるような文に書き換える工程を定式化することを検討している。書き換え工程を定式化することで、書

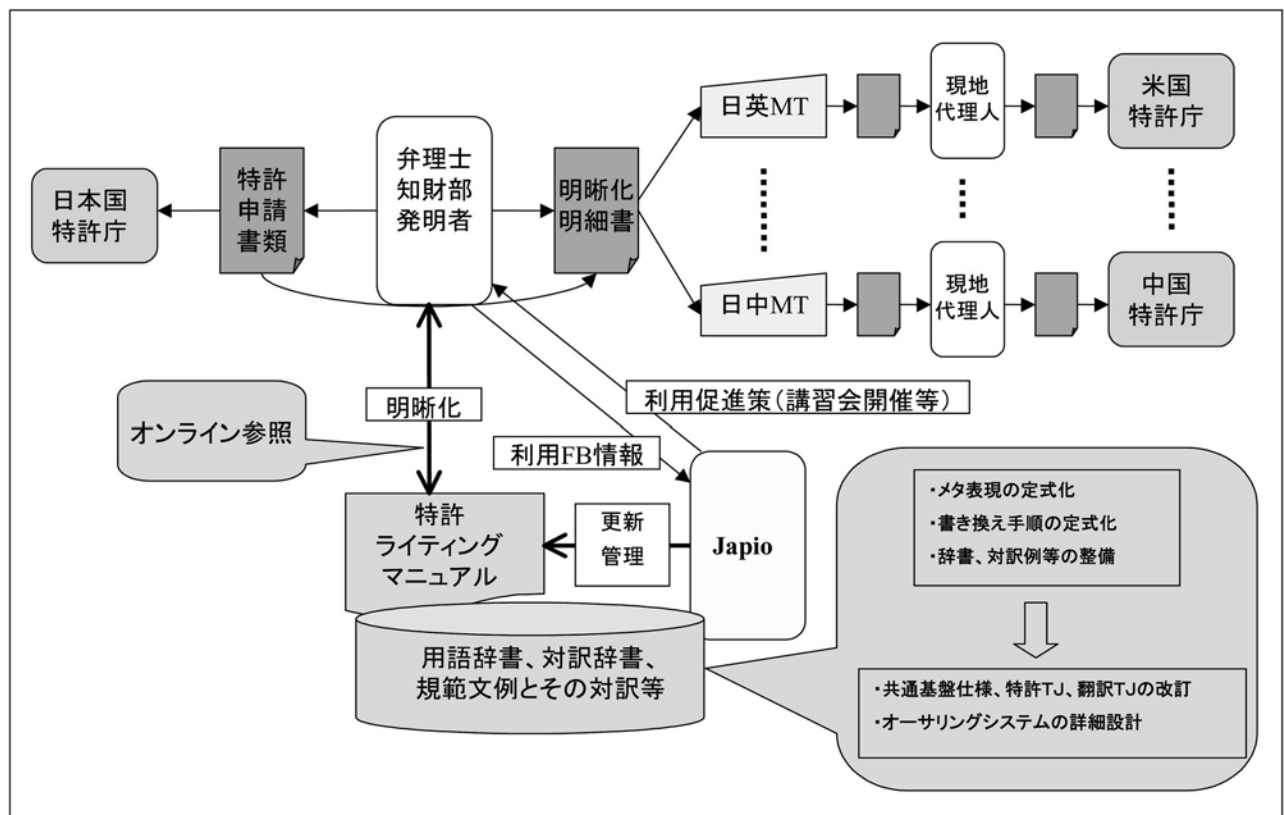


図2 特許明細書ライティングマニュアルの作成・運用

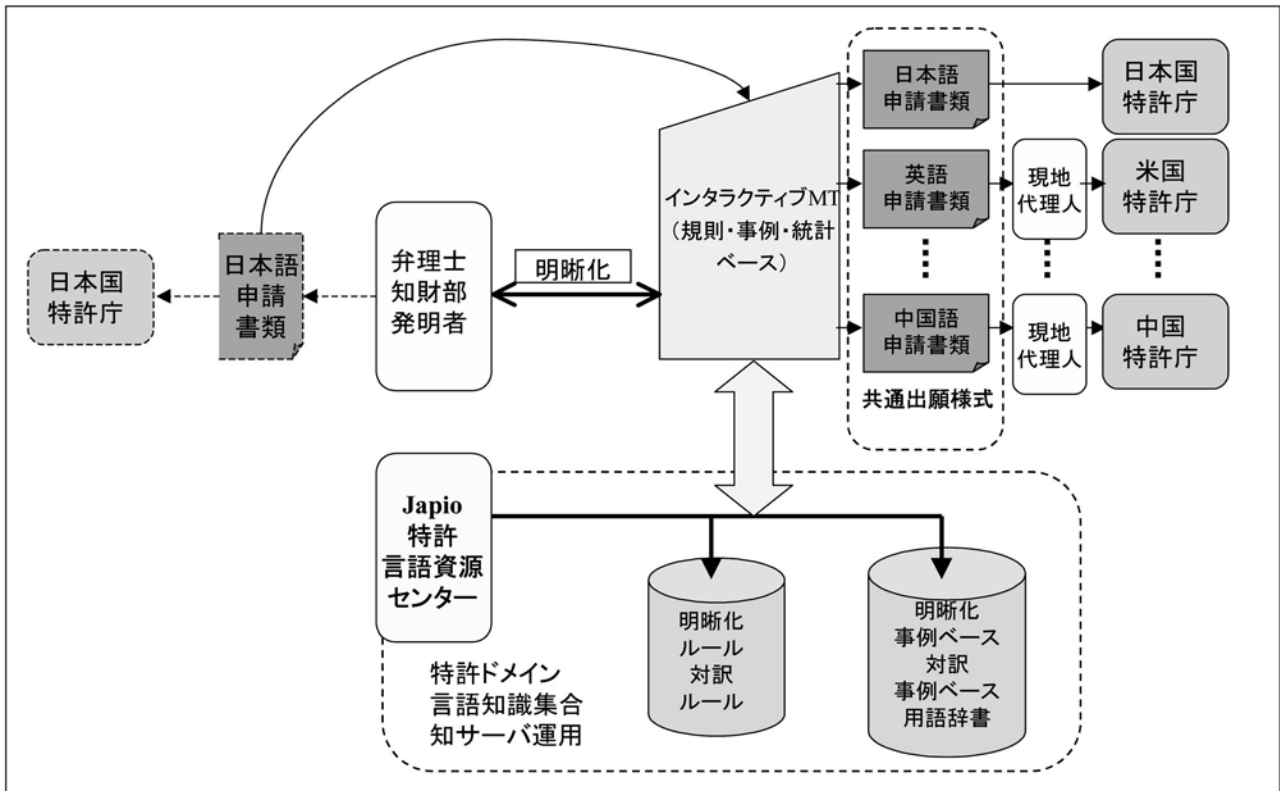


図3 特許版産業日本語プラットフォーム（出願・申請系）の運用

書き換え作業の効率化が図られるとともに、オーサリングシステムの詳細設計時に必要なオーサリング工程を明らかにすることができる。さらに、その書き換えを事例に即して具体的に提示する形で整理したいと考えている。

マニュアル作りを通して得られた情報や知見は、産業日本語共通基盤仕様、特許版産業日本語仕様の検討や改訂作業にフィードバックされる。

一連の作業は、特許版産業日本語委員会において、知財、自然言語処理、機械翻訳システム開発のエキスパートが集い、協調・協力しながら進められる。例えば、言語学的に明晰化された書き換え文が提示されたとき、知財の専門家により、書き換え前後の技術的意味内容や権利範囲の相違の有無等が確認され、機械翻訳システム開発の専門家により、書き換え文の機械翻訳結果の確認及び評価や、機械翻訳向けの再書き換えの検討、現行の機械翻訳の問題点の洗い出し等が行われることで、「技術的な意味の相違や、権利範囲の変動がない書き換えの定

式化」等の有意な作業を効率的に進めることができる。

こうした作業は、多くの人手と労力を要するため、平成21年度は、まず、特定の技術分野を対象に作業を行い、作業遂行上の課題を一つずつ解決しながら、次年度以降、対象の技術分野を拡げていく方向で考えている。

将来的には、作成された特許明細書ライティングマニュアルや、マニュアル作成時に用いた事例や用語集等をデータベース化して、オンラインで参照できる環境を提供するとともに、参照したユーザからの利用状況や改善点等のフィードバック情報が得られる環境を構築していきたいと考えている。こういった環境が構築されることにより、ライティングマニュアルの利便性や運用性のより一層の向上が期待される。（図2を参照。）。

このように、特許明細書ライティングマニュアルの作成や運用を行うことで、オーサリングシステムの詳細設計に必要な情報を集め、また、マニュアル類の更新管理を集合知により行うノウハウを蓄積し、特許出願・申請

系の特許版産業日本語プラットフォームの構築に結び付けることを検討している。

平成21年1月より、日米欧の三極特許庁は、共通出願様式による特許出願の受付を開始した。やがて、中国特許庁や韓国特許庁をはじめ、その他の国の特許庁もこの共通出願様式による特許出願の受付を開始するようになることが期待されている。

特許版産業日本語プラットフォームが運用されるようになった暁には、特許版産業日本語で記述され、共通出願様式に則った、ただ一つの明細書を準備しさえすれば、我が国の特許庁はもちろん、外国の特許庁へも機械翻訳を活用し、今よりもぐっと低い翻訳コストで出願することが可能となるときがやってくるであろう。（図3を参照）

## 4 おわりに

我が国の国際競争力の維持、向上のためには、創造された知財を、国内外で確実に保護し、これを無駄なく活用していくことが肝要である。特に、外国での特許取得をより促進するためには、機械翻訳の有効活用を図る必要がある。また、特許情報の利用性をより一層向上させることが、新たな知財の創造を促し、オープン・イノベーションを加速することに繋がる。産業日本語プロジェクトを推進し、特許版産業日本語を開発・普及することは、これらを実現する上で、極めて重要である。Japioは、経済産業省・特許庁等の関係各省庁、大学や研究機関、ユーザ企業・団体、サービス企業・団体といった関係各位のご支援・ご協力を賜りながら、産業日本語プロジェクトを成功させ、我が国産業界の更なる発展に寄与することを願っている。

平成19年度、平成20年度の「経済活性化のための技術用日本語プラットフォームの開発に関するフィージビリティスタディ」事業は、(財)JK Aの機械工業振興事業補助金の交付を受けて行う(財)機械システム振興協会の委託事業により実施いたしました。