

# 産業日本語とそのプラットフォーム開発計画

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

奥 直也

## PROFILE

1986年 特許庁入庁、2000年 世界知的所有権機関(WIPO)PCT法律部カウンセラー、2003年 特許庁審判部審判官、2004年 電子情報管理室長、2005年 特許審査第二部サービス・安全室長 2006年10月 (財)日本特許情報機構



## 1 はじめに

財団法人日本特許情報機構(Japio)特許情報研究所では昨年度より、産業日本語プラットフォーム開発のスタートを目指して活動を行ってきている。

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表現するための日本語で、Japioの目標は産業日本語プラットフォームの開発と普及である。

本稿では、本プロジェクトの必要性と重要性を説明し、その開発計画の現時点での案を紹介する。

る文書まで、「産業活動」に関わる情報を挙げていけば切りがなく、「産業情報」は、われわれの産業活動に関わる全ての情報と言える。そして日本人が情報のやりとりを日本語で行う以上、日本の産業活動は日本語の上に成り立っている。

しかし現状において、効果的な日本語の利用に関し業界横断的な本格的な規格作りの動きは見られず、技術情報の発信において、国際化も進まない状況にある。日本語をこのままの状況で放置しておけば、英語圏との情報ギャップが拡大する一方である。

Japioでは、こうした状況を受け、この産業に係る日本語に大きな変革をもたらす活動を進めている。

## 2 産業日本語導入の対象となる情報

産業日本語の導入対象となる情報、すなわち、「産業情報」と言えば、まず特許情報、論文情報、営業秘密文書、設計や製造の仕様書、製品マニュアルなどが挙げられ、それだけでも産業日本語導入のインパクトは相当大きなものであろう。しかし、「産業情報」、すなわち産業活動のなかで使われる情報は、そうした、ある程度まとまったものばかりではない。皆様それぞれの仕事の場面を思い浮かべていただければお気づきいただけると思うが、技術を創造し、保護し、また事業として利用し、それらのステップを循環させていく産業活動のなかで飛び交う文書は、千差万別かつ膨大である。会議、発注、説明資料、視点を変えて、社内文書から社外と行き来す

## 3 産業日本語とは

産業情報をよりフレキシブルで扱いやすいものとし、産業活動そのものを円滑にすることで、経済の活性化を促すのが産業日本語である。特に、以下の3点において、その有効性は顕著に表われる。

### (1) 翻訳の品質と効率の向上

原文に着目し、機械処理に適した「産業日本語」をベースに作成されたドキュメントは、翻訳のための機械処理にも乗りやすく、我が国から世界への情報の発信力は飛躍的に向上することになる。

### (2) 業界横断的な情報共有の促進

人に理解し易く機械にも処理し易い産業日本語をベースに作成された情報は、産業界と学界といった情報の垣

根を無くし、更に産業界の異業種間での情報の相互利用も大いに促進することになる。

### (3) 先進的知識マネジメントの確立

産業日本語を更に積極的に適用し、文書自体が知識として共有可能な形態で記述されることになれば、従来の、技術は人に帰するものであり、測れないものであるといった観念が打ち破られることになる。

## 4 産業日本語と知的財産政策

産業日本語は、特に知財との関わりが深い。これを8月に発表された「イノベーションと知財政策に関する研究会」（特許庁）での提言に照らして整理する。

### (1) 「仮想的な世界特許庁」の構築（提言1）

各国特許庁が他庁の審査結果を利用する「審査のワークシェアリング」と「途上国への情報発信」では日本の特許庁審査官の作成する拒絶理由通知書等の内容が機械翻訳され、他庁に発信されるが、その質の向上は不可欠である。そのためには機械翻訳のシステム側への対応に加えて、更に一步踏み込み、元の文書作成に対しても、何らかの取り組みが必要であると考えられる。この仮想的な世界特許庁の実現に関して、研究会の報告書のなかで、Microsoft 社が、「公表された言語に関わらず、先行技術を読み、理解する能力（自動翻訳ツール）が求められる・・・」とパブリックコメントしており、外国において機械翻訳された日本の特許文献が読まれることを想定した、この指摘に対応するためには、特許請求の範囲、明細書等の機械翻訳のより一層の質の向上が求められる。しかし、難解な特許文書では、システム側での対応に限界があり、特許文書の書き方に対して、何らかの取り組みが必要である。その具体策が、産業日本語の導入になる。

### (2) 「オープンイノベーションの進展」（提言7）

「オープンイノベーションの下での知的財産権は、知識・技術の流通を円滑化するためのインフラとしての役

割が求められる。」とある。しかし、難解な特許文書では円滑な知識の流通に適しているとは到底言えない。特許文書の書き方に関して、目的に合った表現の導入などの取り組みが必要である。特に、より高いレベルでの知識の共有を求めるのであればハイパーテキストレベルでのファイルの共有では不十分で、テキストレベルではなく特許情報を文章レベルで捉え、意味に踏み込んで検索や文章理解を可能にし、ドキュメント自体を知識として共有する知識マネジメントが求められる。

### (3) 「シームレスな検索環境の実現」（提言10）

特許情報と技術情報をシームレスに検索できる環境の整備に関して、難解な特許文書が混在するDBでは効率的な検索に支障が生じることが懸念される。ある程度の体系的な対応は可能であろうが、より効率的なシームレスな検索の実現のためには、特許情報とその他多くの技術情報との表現ぶりに共通性を持たせ、(2)と同様、ハイパーテキストレベルでのデータの共有に留めず、それらの内容を文章レベルで捉え意味に踏み込んで検索や文章理解を可能にし、ドキュメント自体を知識として共有する知識マネジメントが求められることになる。

知財政策において提言として掲げられた重要な取り組みに対して、産業日本語の果たす役割は極めて大きく、これらの政策の円滑でかつ効果的な実現に向け、産業日本語は、大いに貢献するものである。

## 5 産業日本語プラットフォームとは

産業日本語は、日本の産業技術文書を人に理解しやすく、機械にも処理しやすいものにすることによって、日本の産業文書が直面している3つの障壁、すなわち「日本と海外の間の障壁」「産業界における分野間の障壁」「人間とコンピュータの間の障壁」を取り払い、その情報源としての発信力を向上させ、ひいては我が国の知的生産性、国際競争力の強化を促すものである。これまでこうした観点から日本語に対して数々の提言や取り組み

みが成されてきた。しかし、その実現には至っていない。理由としては、こうした産業日本語を我が国の産業界に根付かせるには、その導入に取り掛かれるような基盤を開発整備し産業界に提供する必要があるにもかかわらず、いまだ確実な技術に裏付けられた方式とシステムに到達できていないことが考えられる。

Japioが目指す産業日本語プラットフォームは、新たな道を切り開き、産業界が産業日本語という新しい情報化に乗り込みやすくするための基盤となるシステムを提供するものである。ところで、産業技術文書には様々なカテゴリがある。産業日本語をカテゴリのそれぞれに適用していくためには、その文書のカテゴリに応じて個別に詳細な仕様を設定し、システムを開発していく必要がある。産業日本語プラットフォームはその名のとおりに各産業文書カテゴリにおいて、こうしたカスタマイズが行われる際に、その基幹をなす部分として共通して利用可能なオープンプラットフォームとなる。具体的には、産業日本語の中核となる「プラットフォーム仕様」と、産業日本語を作成、利用するための各種「システム」にて構成される。

## 6 産業日本語プラットフォーム 開発計画実施案

平成19年度、Japioは、財団法人機械システム振興協会から開発に関するフィージビリティスタディを受託し、開発課題を整理し開発スケジュールと開発体制を含む開発計画の基本案を策定した。

本年度は、プラットフォーム開発計画の基本案を詳細化した実施案を策定する。現時点での概要は、以下のとおりである。

産業日本語プラットフォームは、中核となる「産業日本語プラットフォーム仕様」、産業日本語で技術文書を作成するための基盤システムである「プラットフォームシステム」、及び産業日本語にて作成される文書を有効

活用するための「アプリケーションシステム」で構成される。【図1】

これらは産業日本語を導入する際の基幹技術としてあらゆる産業技術文書カテゴリにて共通的に利用可能なオープンプラットフォームである。

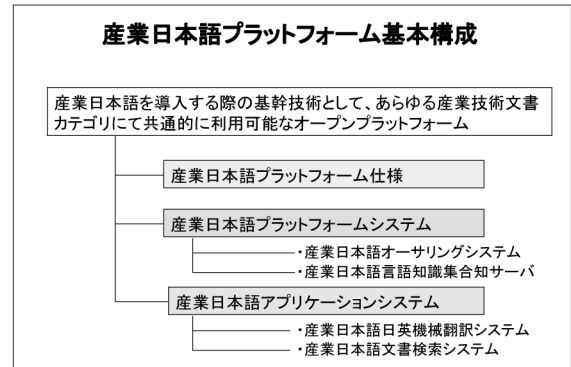


図1 産業日本語プラットフォームの開発課題

### 6.1 産業日本語プラットフォーム仕様

産業日本語は、種々の目的に応じて多様な仕様の設定を許容する。「産業日本語プラットフォーム仕様」は、これら目的に応じて設定された種々の詳細仕様を生成するベースとなる。本年度のスタディでは、現在の暫定版を、第1版に改訂する予定である。

そして、プラットフォーム開発期間は、他の構成要素である「プラットフォームシステム」及び「アプリケーションシステム」の開発のベースとしつつ、開発実施の過程で生じた課題や問題点を反映させて、更なる改訂を行うことになる。

### 6.2 プラットフォームシステム

策定した産業日本語プラットフォーム仕様を我が国産業界に迅速かつ広範に普及させるために次になすべきことは、実際に産業技術文書を作成する現場に対し産業日本語を用いた文書作成を支援する技術を提供し、予め作成者の負担を軽減する環境を整備しておくことであって、プラットフォームシステムは、この目的のために構築される。

プラットフォームシステムは産業日本語による文書作

成を支援するワードプロセッサ的機能を有する「産業日本語オーサリングシステム」と、産業日本語を用いた表現や非明晰な日本語を明晰な産業日本語に変換する事例等を蓄積した辞書データベースを格納した「産業日本語言語知識集合知サーバ」により構成される。そして、これらのシステムの実現のための基幹技術として、ISeC (Institute of Semantic Computing, 特定非営利活動法人セマンティック・コンピューティング研究開発機構) が開発したセマンティック・コンピューティング研究開発推進のための共通言語CDL (CDL.jp.sf: 日本語表層概念記述用CDL) を用いる。

### 6.3 アプリケーションシステム

産業日本語による各種技術文書の作成は、読み手すなわち人間に対する明晰性の向上と同時に、コンピュータ処理に対する高い適合性も実現するものであって、この特性を最大限利用し産業日本語文書の入力を前提とした各種アプリケーションシステムが開発・提供されることで、各種産業技術文書の利用性は飛躍的に高まり、我が国の知的生産性の向上につながることになる。

開発計画実施案では、主要なアプリケーションシステムの例として、「産業日本語日英機械翻訳システム」及び「産業日本語文書検索システム」の開発を実施することとしている。いずれも産業日本語文書の有するCDL情報を最大限に活用することで、より高い次元での文書処理を達成するものである。

### 6.4 開発スケジュール

現在の計画では平成24年度からの運用を目指し、平成21年度から平成23年度までの3ヶ年を開発フェーズとしている。

開発するシステムの具体的な内容については、協力・連携先のニーズを踏まえ各種技術文書にて共用できる根幹技術に基づきつつも、特定の技術文書を対象に速やかに実用化が可能な形とする必要がある。すなわち、プラットフォームとして開発した根幹技術を、特定の技術文

書の創成・交流・活用サイクルに導入した「モデル運用サービス」についても、実用化する必要がある。

Japioでは、明晰化の必要性、対象となる文献の量、検索や翻訳の重要性などから、特許文書が産業日本語化のメリットがとりわけ大きいジャンルであると考え、特許庁をはじめとする関係各機関との連携のもと、まずは特許文書をテーマとしたモデル運用サービスの開発について検討を進めていきたい、と考えている。

### 6.5 産業日本語普及のための体制

この開発を成功させ、その先の運用フェーズへの移行をスムーズなものとし、産業日本語の我が国産業界への速やかな普及を実現させるためには、主要なベンダー企業、ユーザ企業、関係団体及び関係省庁とが意識を共有し、一致団結して取り組んでいくことが何よりも重要である。Japioではそのための体制として、産業日本語コンソーシアム及び産業日本語フォーラムの立ち上げを検討している。

産業日本語コンソーシアムでは、ユーザ側、サービス側の企業・団体を中心に、サービス事業の立ち上げや普及を目的に活動し、産業日本語フォーラムでは、大学や研究機関を中心に、産業日本語に対する学術的な裏付けや、プラットフォームシステムを研究開発ツールとして活用する新技術研究開発の促進等の活動を推進していただくことが狙いである。

## 7 産業日本語プラットフォームの運用イメージ

最後に、今後どのような世界が産業日本語によって拡がることになるのかを、産業日本語を特許文書に適用した際のイメージで説明する。

特許文書について指摘される問題点としては、まずは一文が長く、かつ係り受けが複雑で読み手にとって分かりづらいという点が挙げられる。また、用語の統制ができておらず、同じ用語がさまざまな表記や同義語にて表

現されているということも、その利用性を低くしている。これらの問題点は、読み手に正確に内容が伝わらないという根本的な損失をはじめ、その難解さに起因する特許庁における審査効率への悪影響、先行技術検索のしにくさによる重複開発などの不効率、さらには機械翻訳への適合性の低さによる過大な翻訳負担を科し外国出願へのインセンティブを低下させ、翻訳ミスによる海外での権利取得の失敗など、我が国産業界の生産性向上に数多くの不都合を及ぼしている。【図2】

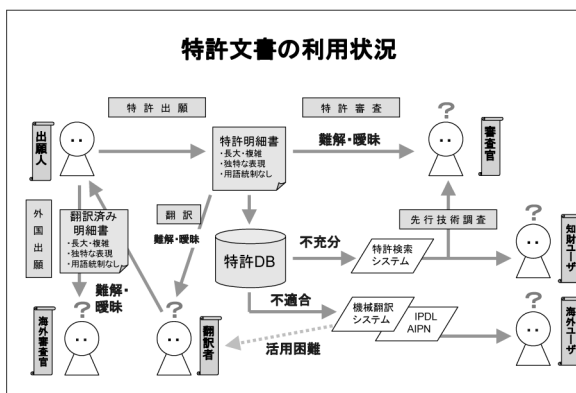


図2 現状の特許文書の問題点

産業日本語プラットフォームは、これらの問題点を大きく改善するもので、特許文献の作成にあたり、長すぎる文や難解な係り受け、誤解を生む表現をアラートし、適正な表現による記述を促すインタラクティブなオーサリング技術を提供する。

これに適用するルールは、Japioが現在策定中の特許版の産業日本語の仕様に基づく。そして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として、階層化された集合知データベースを構築する。格納された情報は、オーサリングシステムと連携しその内容に応じて企業内や産業界全体で共有され、協調作成の形でアップデートされていく。開発計画では、このデータベースの共通仕様を策定し、階層に応じた管理と社内や産業界全体でのWebを介した共有を可能とするためのインタフェースやサーバの開発を行うこととしている。【図3】

これらの技術を、明細書作成作業や、その前段階にお

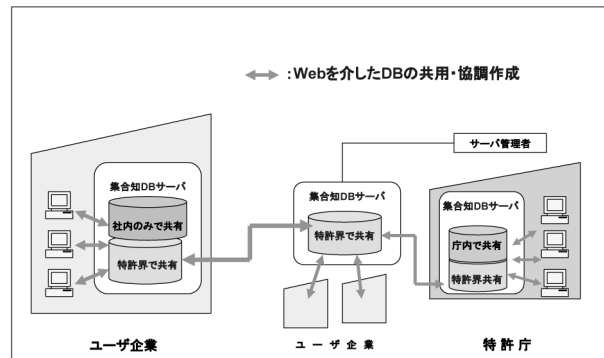


図3 集合知データベース/サーバ概念図

ける技術文書作成作業に導入することで、出願人は最小限の負担にて自らの特許出願明細書を産業日本語で記述できるようになる。こうして産業日本語化された特許文書は、その明解かつ統制された内容により、特許庁における審査効率の向上をはじめ、翻訳や検索への適合性も優れていることから、翻訳精度・検索精度の大幅な向上が見込める。更に、産業日本語オーサリングシステムを介して作成された文書は、半自動的にCDL化される仕組みにしようと考えており、これによりCDL技術を最大限に活用した、産業日本語文書専用の機械翻訳システムや高精度の検索システムが実現可能となるであろう。このような高精度の機械翻訳システムの実現によって、明晰化された産業日本語文書との相乗効果により、現在は全面的に高価な上級翻訳者に頼っているといえる海外出願のための明細書翻訳も、かなりの部分が正確に自動翻訳できることとなり、翻訳費用の軽減、ひいては我が国の外国出願のインセンティブの向上につながるものと考えられる。また、この技術は日本から海外へ発信する特許情報の品質も大いに向上させるであろう。

検索システムに関しても、CDLの導入により、文書を構成する各語の係り受け関係や階層化された概念に基づく精細な検索が実現し、特許審査や民間における先行技術調査などの大幅な精度向上と効率アップが達成される。【図4】

そして更にこの先には、過去分の特許文献を自動的に産業日本語化するツールの開発や、特許文書にて実現し

たプラットフォームの、他の分野への展開も考えられる。

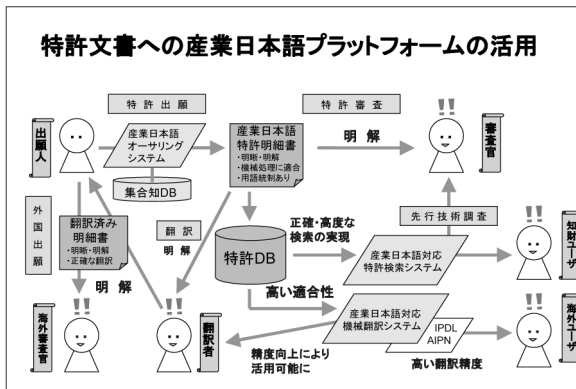


図4 産業日本語プラットフォームの活用イメージ

## 8 おわりに

グローバル経済下で、外国での特許取得が不可欠であることは言うまでもない。しかし、出願明細書を限られた期間内に正確に翻訳するための大きな負荷と翻訳品質の不足が我が国の外国出願・特許取得の際の大きな障壁となっている。産業日本語は、この障壁を取り払い、我が国知財のグローバル化を加速し、国際競争力を迅速に強化するものである。

他方、特許情報は最も体系的に整理され、極めて有用な技術情報であるにもかかわらず、慣習的に用いられる独特の表現形態や記載様式により、読むのも機械で処理するのも困難である。この状況に対し、産業日本語を導入することで、特許情報の技術情報としての活用が一気に加速し、出願公開された技術が多くの研究開発者に的確に伝わることで、二重の投資が回避され、次なる発明のための基礎的な技術情報として大いに活用されることになる。

産業日本語の導入によって、特許制度の本来の機能がより有効に発揮されることになり、産業日本語は、真の知財立国の世界を拓げていくことになる。

