



特許庁
特許技監

南 孝一



独立行政法人工業所有権情報
・研修館理事長

清水 勇



NTTアドバンステクノロジー株式会社
取締役/知的財産ビジネスセンタ所長

澤井 敬史



独立行政法人産業技術総合研究所
サービス工学研究センター次長

橋田 浩一

H20.8/1~H20.9/30

産業日本語特集 ネット座談会

特許版産業日本語の取り組みと期待

— 特許情報を高価値化する産業日本語 —



東京大学情報理工学系研究科教授/国際
言語処理学会(ACL)前会長(2006年)/
アジア太平洋機械翻訳協会前会長/アジ
ア言語処理学会連合(AFNLP)会長

辻井 潤一



株式会社東芝 研究開発センター
知識メディアラボラトリー室長

木下 聡



財団法人日本特許情報機構
専務理事

守屋 敏道



財団法人日本特許情報機構
特許情報研究所顧問

横井 俊夫

はじめに 産業日本語の役割と位置付け

進行 横井 : 8月1日(金) 投稿スタート

座談会の進行役を担当いたしますJapio特許情報研究所顧問の横井です。私自身は、永らく、ITの基礎・基盤技術の研究開発、とくに、言語や知識に関する研究開発に携わって参りました。そして、今、痛感しておりますのは、これからのITの新技术開発には、情報そのものの利用やサービスからアプローチすることが非常に重要になってきたということです。情報の利用やサービスといえますと、**情報表現の根幹を担う日本語、特許情報における利用やサービスに関する蓄積**、この2つは、非常に重要な手掛かりとなります。この2つの手掛かりが結び合わさるところに産業日本語という設定が生まれました。

昨年度は、東京工科大学メディア学部教授の立場で、Japio特許情報研究所のいくつかの委員会に協力させていただき、特許情報の現状とこれからの課題について親しく接する機会を得ました。そして、特許情報自身が、さらなる高価値化に向けIT新技术を希求している有様を実感し、また、ITの技術革新に向け特許情報が非常に良

い素材となることを実感することができました。

今年のYEAR BOOKは、「産業日本語－特許情報を高価値化する産業日本語」に関する特集号となります。専門家の方々のご意見は、寄稿集をご覧頂くとして、このネット座談会では、産業日本語の役割をより広い立場からご議論いただき位置付けていただくこととなります。議論を効果的に進めるために、まず、守屋Japio特許情報研究所長にJapio特許情報研究所における産業日本語に関わる取組みをご説明いただきます。その説明を軸にして、それぞれの座談会メンバーの方々にポジションペーパーのような形式で同時並行にてご意見を頂きたいと思っております。頂いたご意見を加味して論点を整理いたしますので、さらに、その論点を巡ってご議論頂くという手順を進めたいと思っております。

それでは、まず、『産業日本語プラットフォーム開発計画』と題して、守屋所長にJapio特許情報研究所の取組みをご紹介します。

産業日本語プラットフォーム開発計画

守屋 : 8月4日(月) 投稿

昨年(2007年)のYEAR BOOK－特許情報研究所設立記念版－のネット座談会は、「特許情報に関する幅広い専門家の交流の場としての役割を果たすこと」「今後の我が国の特許情報研究のキーステーションとして大きく発展すること」というJapio特許情報研究所への強い期待で締めくくられています。この期待に応えるため、特許情報研究所が初めての本格的なプロジェクトとして進めておりますのが、『産業日本語プラットフォーム開発計画』です。

これまで、客観的な技術情報を表現・伝達する産業技術文書において、日本語を明晰に使用するための工夫や提言は、無数といわれるほどに行われてきました。しかし、いまだ、確実な技術に裏付けられた方式やシステムには行き着いてはおりません。このいまだ無かった日本語の明晰な用法に関する方式とシステムを実現しようというのが産業日本語プラットフォーム開発の目標です。

産業日本語とは、一言でいいますと、日本の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するための日本語です。産業日本語を我が国の産業技術文書に広く導入し、我が国の産業技術情報の情報力を飛躍させれば、我が国の知的生産性の更なる向上と経済活性化が期待できると考えております。

なぜ、今、このような目標設定が可能であるのかについて、技術的な要件としては、次の3点が挙げられます。

(1) 日本語に関する言語学的・言語処理的な知見や経験

P R O F I L E

横井 俊夫

財団法人日本特許情報機構
特許情報研究所顧問
1966年に電気試験所(現在:産業技術総合研究所)。1982年より第五世代コンピュータプロジェクトの推進に従事。1987年より電子化辞書プロジェクトに従事。1995年よりフィリピン国にてODAプロジェクトの推進、指導に従事。1997年より東京工科大学、2008年より現職。
特許版産業日本語委員会副委員長



の蓄積が十分になされた。

- (2) ワープロ技術や校正・推敲支援技術、そして、機械翻訳技術などの処理技術が成熟し、言語資源等の整備やWeb環境における言語知識の集合知の仕組みなど、システム実現技術が適切に整ってきた。
- (3) 翻訳・検索・要約等の文書処理を様々な方式の知識処理やマルチメディア処理と連携処理させるための共通言語CDL (Concept Description Language : 概念記述言語) が、プラットフォーム開発の根幹技術として導入できる状況に達した。

これまで、これらの蓄積は個別で、ばらばらの試みの累積でした。産業日本語プラットフォームの開発では、こうした我が国のアドバンテージを、体系的な方式やシステムにまとめ上げることとなります。

さて、これまでの活動の経緯をご説明しますと、開発計画では、2007年度と2008年度をスタディフェーズと位置づけ、まず2007年度に「特許版・明晰日本語策定委員会」を立ち上げ、特許文書を明晰に記述するための日本



守屋 敏道

財団法人日本特許情報機構
専務理事

昭和49年特許庁入庁、平成
9年総務部国際課長、平成
11年総務部特許情報課長、
平成13年審判部審判長、特
許審査第一部調整課長、平成
15年特許審査第三部長、平
成16年審判部長、平成17
年特許庁特許技監、平成20
年7月より現職



語のあり方を広く調査し、「明晰日本語基本仕様(第0版)」を策定いたしました。これは、特許分野の専門家と言語分野の専門家が初めて、本格的な共同作業を行うという大変意義深いものとなりました。

そして、2007年度後期には、財団法人機械システム振興協会から「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ェージビリティスタディ」を受託し、特許明細書の明晰化実験を実施し、人間に対する可読性の向上と日英機械翻訳システムに対する翻訳精度の向上の2つの観点から評価を行ない、技術用日本語の有効性を実証しました。そして、「開発計画の基本案」を策定いたしました。

今年度(2008年度)は、この受託事業を継続し、実験用ソフトを作成し、システム開発のフェージビリティを実証し、「開発計画の実施案」を策定いたします。他方、明晰日本語基本仕様は、システム開発を本格的にスタートできる精度の仕様(第1版)を策定いたします。

ところで、これまでの説明の中で、事業立ち上げの経緯もあって、「産業日本語」「明晰日本語」「技術用日本語」という3つの用語が入り乱れておりますが、本プロジェクトを効果的にPRし、プロジェクトについて共通の認識を持っていただくために、去る7月18日に開催したワークショップから「産業日本語」に統一いたしました。この「産業日本語ワークショップー産業情報基盤確立のための明晰な技術用日本語である産業日本語ー」には、多数の関係者にお集まりいただき、非常に熱心なご議論をしていただき、共通認識の確立、関係部署へのアピール、活動母体の拡大という目的を達成すること

ができました。以上が、これまでの活動の経緯です。

それでは、次に、来年度以降の開発計画を、課題、スケジュール、及び体制に分けて簡単に説明いたします。
『開発課題』は、大きく、「プラットフォーム仕様」、「プラットフォームシステム」、「アプリケーションシステム」、および「モデル運用サービス」の4つから成り立っています。

- (1)「産業日本語プラットフォーム仕様」は、様々な産業日本語に対して、明晰性を体系的に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からなります。産業日本語はそれぞれの用途に対応して様々な日本語仕様となります。
- (2)「プラットフォームシステム」は、オーサリングシステムと言語知識集合知サーバからなります。オーサリングシステムは、明晰な日本語文書の作成をインタラクティブに支援し、規則ベースや辞書ベースの内容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出来ます。他方、集合知によって、この規則ベースや辞書ベースの内容を漸進的に整備していく共通環境が、言語知識集合知サーバです。
- (3)「アプリケーションシステム」では、文書処理の代表として翻訳と検索を取り上げます。産業日本語によって機械翻訳や文書検索の高度化が効果的に達成されることを実証します。
- (4)「モデル運用サービス」では、特許関連文書を対象にした文書の制作、蓄積・検索、変換(翻訳や要約)を高度に支援するワンストップサービスと、企業内文書と企業内知識オントロジーに基づく先進的な知識マネジメントサービスとを取り上げます。(詳しくは、寄稿集にある「産業日本語プラットフォーム開発計画」をご覧ください。)



次に、『開発スケジュール』ですが、大きくスタディフェーズ（2年間）、開発フェーズ（3年間）、運用フェーズの3段階からなり、平成24年度以降を目途とする運用フェーズでは、後述の産業日本語コンソーシアムを母体として、各種サービスの立ち上げ促進と相互連携促進を進め、産業日本語をベースとする知の高度生産環境の統合的な整備への活動を展開することを想定しております。

そして、そのための『開発体制』については、プロジェクトマネジメントを、Japio特許情報研究所が中心となって行い、開発の実施検討にあたっては、東芝、富士通、日立、ジャストシステム、数理先端技術研究所等のベンダー企業、三菱総研、トヨタ、キヤノン等のユーザ企業、そして、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等に参加していただく予定です。

平成21年度には、『産業日本語コンソーシアム』と『産業日本語フォーラム』を立ち上げることができればと考えております。『産業日本語コンソーシアム』は、ユーザ企業・団体やサービス企業・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するものです。また、『産業日本語フォーラム』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念頭においています。

Japio特許情報研究所としましては、まずは産業日本語プラットフォームの開発を成功させることが当面の課題となりますが、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するためには、産業界の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あらかじめ意識を共有し、一致して取り組んでいくことが何よりも重要であると考えています。

『産業日本語プラットフォーム開発計画』を巡るご意見

進行 横井 : 8月6日 (水) 投稿

守屋所長に説明いただきました『産業日本語プラットフォーム開発計画』を巡る形で、まずは、それぞれの方々からご意見をいただきたいと思います。

南特許技監には、これからの特許行政とこれからの特許情報のあり方、ならびに、これからの特許関連文書（特許明細書等や特許審査文書）のあるべき姿をお話いただき、産業日本語プラットフォームを位置付けていただきたいと思います。

清水理事長には、多くの課題がある大学や研究機関に

おける知財活動の活性化、特許情報と学術情報の連携等々を背景にこれからの特許明細書等のあり方を論じていただき、その中に産業日本語や開発計画を位置付けていただきたいと思います。

澤井取締役には、これからの日本企業の知財戦略に与っての特許情報のあり方を論じていただき、企業における特許明細書等明晰化への取組みと産業日本語や開発計画の位置付けを論じていただきたいと思います。

辻井教授には、これからのIT関連の新しい技術動向の中で産業日本語を論じていただきたいと思います。言語処理技術の動向や機械翻訳技術の動向の中に産業日本語や開発計画を位置付けていただきたいと思います。

橋田次長には、産業技術としてITを革新する立場から、産業日本語を論じていただきたいと思います。サービス工学の観点、サービス工学からみた特許情報サービスの観点から産業日本語や開発計画を位置付けていただきたいと思います。

木下室長には、産業日本語プラットフォームを開発するという側面から、産業日本語を論じていただきたいと思います。機械翻訳ソフトをはじめとするITビジネスや、さらに、日本語処理ビジネスを展開する立場から産業日本語や開発計画を位置付けていただきたいと思います。

『産業日本語』への期待

清水 : 8月18日 (月) 投稿

『産業日本語プラットフォーム開発計画』なるものが粛々と進められ、平成24年度の運用フェーズまでのロードマップまで準備されていることを知って感激しました。この開発計画に期待を寄せる理由として2点あげさせていただきます、私の基本スタンスといたします。



清水 勇

独立行政法人工業所有権情報・研修館理事長
 1969年東工大大学院理工学博士課程修了、85年同大学教授、98年同大学フロンティア創造共同研究センター長、99年理工学振興会（東工大TLO）常務理事、2000年同大名誉教授、理工学振興会専務理事、2004年から現職。



(その1)

グローバルな知識経済社会への変革を目指す我が国において、科学技術研究開発の成果は単なる人類共有の知識を創成するだけでなく、知的財産として広く世界で活用し、我が国に経済的な繁栄をもたらすための知識資源と捉えられています。したがって、大学の研究者といえども、好むと好まざるとその研究成果は、特許等の知的財産として確保することが期待されます。

しかし、私をはじめ特許の明細書を読んだ時を思い出すと、書式は兎も角、文章自体が母国語で書かれているにも拘わらず、故意に難解になるように工夫されているようで、悪意すら感じました。最近、知人の一人が企業の知的財産本部の最高責任者に任命された際、自社の出願特許の文章があまりにも難解なので、部長命令として「2行より長い文章を書かないこと」を提案したところ、猛烈な反対に遭遇し、妥協点として「限界を3行とする」という笑い話を聞いたことがあります。

科学技術に関する知識を持った人にとっても、特許等の文章は他の科学技術論文のそれに比較して明らかに異

質なものを感じます。一方、ライフサイエンス等のグローバルな研究開発競争の激しい分野の研究者にとって、学术论文と共に特許情報は、日常的に触れなければ競争に勝利することが出来ない時代になりました。したがって、学术论文までとは言いませんが、**特許の文章を作成する際にも著者に対して、“一貫性があり、平易で簡潔・正確な記述をする”ことを求めるべきではないでしょうか。**プロ・イノベーション時代をリードするためには、**知的財産は可能な限り広く世界に活用されてこそ経済的な価値が創生されることを再認識すべきです。**『産業日本語プラットフォーム開発計画』は、機械翻訳による多言語化の実現も視野にいられていて、日本における科学技術論文の世界にも、その実現は大きなインパクトをもたらすことでしょう。

最近、特許庁から公開された「イノベーションと知財政策に関する研究会報告書」によれば、“特許情報とグローバルに存在する科学技術情報をシームレスに検索できる環境整備”が提案されています。『産業日本語』の実現は、この特許庁の提案とリンクすることで、グローバルに活用可能な科学技術情報インフラとなる可能性を秘めています。

(その2)

科学技術に関しては、特異な分野を除けば国際会議等の用語は「英語」であり、論文も「英語」で書くことが一般常識となっています。また、日本の学会の多くは英文誌も発行していて、そのことが学会の財政を少なからず圧迫しています。研究成果の発表を取って母国語でない「英語」を使う理由は簡単です。成果を「英語」で発表しないかぎり、世界の研究者に自分の研究成果が評価されないからです。さらに、研究者の業績評価として論文が使われ、論文の質の尺度の一つとして被引用論文数が用いられます。したがって、高い評価を受けるためには、世界の人々から広く読まれているジャーナルへ投稿

する必要があります。多くの研究者が、“Science”、“Nature”への掲載を目指すのはそのためなのです。勿論、研究業績の評価結果は、研究資金の配分にフィードバックされるので、研究者にとって、「英語」の呪縛から逃れることは困難なのが現状です。

そのようなことから、我が国の理工系大学院の研究・教育現場では、「英語」論文の読み方、書き方に関しては、かなりの時間と労力を費やすこととなります。通常、母国語以外の言語を使って、「全体に統一が取れ、わかり易く簡便な表現」で研究成果を正確に記述するためには、科学技術の理解とは別に「英語」に





よる文章の記述に相当な実践的訓練が必要です。

米国のシカゴ大学に在籍した時の経験ですが、英語を母国語としない留学生で、10年近く米国で生活している学生の学位取得のための最終試験に立ち会った時のことです。その際、審査にあたった教授からきままって要請されるのが、論文の科学的内容でなく、「英語」のリバイスです。また、私の指導の下で、日本語で博士論文を書いた留学生の文章には、日本語で喧嘩も出来るほど達者に日本語を操る有能な学生なのに、書かれた論文には、母国語ではおかさない奇妙な表現の誤りが随所に発見され、学位論文としての「品位」を保つため言葉の修正をしました。言語の障壁、特に「文章として書く」ことを完全マスターすることは母国語でも容易いことではありません。

かつて、英語論文を書くときにお世話になった、参考書「英語の化学論文」の書き出しに、－英語国民は英語の論理構造にしたがってものを考え、日本語国民は日本語の論理構造にしたがってものを考える習慣がついているものであるから、同じ英語の化学論文を書いても、日本語国民が陥りやすい誤りがあり、まずさがある。そこで、学習法としては、英語の論文を読むとき、表現の仕方に気をつけ、メモしておく（英借文のすすめ）－とありました。私も、よく使用する「英語」表現をノートにメモして活用させていただきました。語学に特別な才能がある人か、外国語を習得するための環境に恵まれた人以外、言語での苦労は世界共通です。そこで、科学技術分野のジャーナルに投稿すると、「英語」を母国語とした査読委員が丁寧に文章をチェックし修正箇所等を親切に書き込んでくれるか、「英語」を母国語とする人に「英語」の修正してもらうこと勧めるメモが査読結果に書かれているのが一般的です。

それならば、いっそのこと公用語を「英語」にしたら、との提案も聴きますが、その意見には賛成できません。その理由は、言語は文化の根源であり単に情報伝達のツ

ールとして使われるだけでなく、思考プロセスに深く関与しているからです。本年度の芥川賞作品「時が滲む朝」の著者、ヤン・イー女史は中国人で、日本の大学で日本語を習得し、母国語でない日本語で最高の文学賞を受賞されました。彼女の発言に一私の持っている言語感覚というのは、窓のようなイメージで、新しい言語を習得することで、新たな窓が開かれ、見たこともない風景が目の前に広がるようなものです－とありました。ただ、利便性をもとめて大切な日本の文化を見るための「窓」を閉じるようなことには賛成できません。

他方、世界で年間約100万件刊行される査読付きSTM（科学・技術・医学）論文（5万件は日本から）は、世界に2000社ある出版社の内上位5社（Elsevier, Springer, Taylor & Francis Blackwell, Wiley）から刊行されていて、一極集中がかなり進行していることが分かります。さらに、グローバルな知識経済社会では、IT技術の進歩と共に、アーカイブ、機関レポジトリ、セマンティックウェブ等、利用者が必要に応じてオープン・アクセスできる新しいビジネスモデルが近い将来の姿のようです。このままでは、「英語」一色の世界になるのでしょうか？私は「否」と答えたい。本来、IT技術は、情報を世界に解き放ち、誰でも、どこでも、いつでも情報を獲得できる手段であるが、そのために文化の多様性が消滅するほど人間は「ヤワ」にできていません。

一方、**特許等の知財情報は地属性が高く、日本語で書かれている希少な科学技術情報**です。そこで、『**産業日本語**』により、「日本語」の質が担保され、その上で**機械翻訳を介した多言語化が実現され、学術情報と特許情報がシームレスに検索できるグローバルシステムが整備されれば、少なくとも科学技術分野におけるイノベーションは言語の呪縛から解き放たれると期待**されます。

敢えて言うならば、「英語」に統一された世界に隷属するのではなく、**IT技術と「知恵」を駆使することによって、「日本語」の独自性を守りつつ、世界に遍く情報交換できるシステム構築を目指す『産業日本語』の将来に期待**します。

進行 横井 : 8月18日(月) 投稿

特許文書における難解な文章を改善しなければならぬことから、まず、産業日本語を位置付けていただきました。そして、英語と日本語という問題をご指摘いただきました。科学技術の分野では、国際共通語である英語

への対応能力を高めることは、最重要課題のひとつです。英語への対応能力が、単に英語能力だけの問題ではなく、日本人の思考言語である日本語能力の裏付けに支えられるのであるというご指摘は、産業日本語を位置付ける欠くことのできない視点です。

特許行政と特許情報の観点から

南 : 8月20日 (水) 投稿

守屋所長のおっしゃるように産業日本語を「日本の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するための日本語」と定義するとすれば、現在の特許関連文書（特許明細書および拒絶理由通知書等の特許審査文書）に記載されている日本語は産業日本語からはかなり離れた位置にあるといえます。一例をあげれば、最近はその傾向は緩和されてきているかもしれませんが、以前は権利範囲を規定する特許請求の範囲を極度に抽象概念化したり、曖昧にして記載する出願人も見受けられ、そのため特許庁の審査官は特許請求の範囲に記載されている内容を理解するのに苦労することもしばしばありました。また、特許明細書等においては一文が長い等の特徴があります。そのような状況にあってもなお特許関連文書は、日本の産業上、技術文書として非常に有用なものであることから、その明瞭化が必要ではないか、というのが今回の産業日本語の取り組みにつながっているものと認識しております。

そこで、私からは、特許行政と特許情報という観点でお話しさせて頂きたいと思います。以前より日本は特許出願大国であるということは新聞報道などで皆さん既にご存知のことかと思われませんが、現在、特許出願の増加



に対応すべく審査の効率化を求められているというのは日本だけではありません。例えば、2005年において、世界の特許出願総数の約166万件の約77%に相当する約128万件が日米欧中韓の特許庁に出願されていますが、この五庁に出願されている特許出願の約30%が、そのいずれかの他庁にも出願されているのです。このように国際的に見て多くの重複出願がある状況ですので、本年5月に米国特許商標庁で開催された日米欧中韓の五大特許庁会合において、審査における国際的な協力・ワークシェアリングが重要であるとの共通認識が得られたことは至極当然であるといえます。審査におけるワークシェアリングの原則は、最初に出願を受けつけた国の特許庁（第一庁）が、最先にサーチ・審査結果という特許情報を発信し、他庁（第二庁）がその結果を利用することにより審査を効率化するというものです。



では、具体的にどのような特許情報が、ワークシェアリングを行う上で有用性が高いかといえば、先行技術文献として引用されている特許公報は勿論のこと、担当した審査官がその先行技術文献を用いてどのような論理構成によりその出願を拒絶したのが記載されている拒絶理由通知書があげられます。

このような有用性の高い特許情報を海外に発信すべく、特許庁では1999年から特許公報等をインターネットを通じて無料で検索・閲覧可能な特許電子図書館（IPDL）サービスを公衆向けに提供していますが、特

に英語圏のユーザのために公開特許公報の英文抄録（PAJ）に加えて、特許公報全文を英語に機械翻訳して提供しています。

しかしながら、その特許出願の技術内容を概ね理解する上では有益であるとの評判は頂いているものの、前に述べたとおり、抽象的であったり、曖昧であったり、長文であったりする特許請求の範囲や明細書の機械翻訳は、その技術内容を正確に理解するためには不十分である旨のご意見も頂いています。

また、特許庁では2004年より、海外特許庁に対して高度産業財産ネットワーク（AIPN）と呼ばれる特許公報、サーチ及び審査結果に関する情報を英語に機械翻訳して提供するサービスを行っており、このAIPNは現在30カ国以上の海外特許庁で利用可能となっています。そして、AIPN利用国では日本からの特許出願に対応する日本出願について日本国特許庁での審査結果を参照することによって審査に係る負担を軽減できるとともに、審査の質の向上を図ることができることから、最終的には日本の出願人の海外での適切な権利取得につながっています。

さらに、AIPNは、第一庁で特許となった出願について第二庁において簡易な手続で早期審査が受けられる施策「特許審査ハイウェイ」においても活用されており、結果として海外における日本からの出願に対する権利取得の迅速化につながっています。

しかし、拒絶理由通知書等の審査関連書類は特許明細書で用いられている専門的技術的用語に加えて、法律用語や審査関連文書特有の表現を含むものであるため、時として日本語で読んでも意味が伝わりにくいことがあります。このような場合、当然のことながら質の高い機械翻訳を期待することはできません。

これに関して、審査関連書類の記載内容を含む審査の質全体については、従前より行われていた審査部における複数の審査官による協議や管理職等による内容チェックに加え、さらに、2007年4月に設置された「品質監理室」における審査の結果の事後的、客観的な測定・分析およびその結果の施策反映等により、一層の審査の質の維持・向上に努めているところです。そして、より良い機械翻訳結果の提供という観点から、昨年度は進歩性の拒絶理由を含む拒絶理由通知書について、拒絶理由通知書中の定型文（条文）の変形表現、及び、自由記載部分から定型化可能な表現を分析し、その分析結果に基づき機械翻訳結果の品質向上につながる表現の辞書登録等を行い海外特許庁審査官によるサーチ・審査結果の利便性向上を図りました。

今後、このような機械翻訳辞書の増強だけでなく、機械翻訳の入力データとなる日本語の文章自体が、例えば、提案されている産業日本語プラットフォームの確立

P R O F I L E

南 孝一

特許庁特許技監
昭和52年特許庁入庁、平成13年
総務部特許情報課特許情報利用推
進室長、平成14年総務部技術調
査課長、平成16年特許審査第一
部調整課長、平成18年特許審査
第二部長、平成20年 7月より現
職



により、機械翻訳に適した文章にされることにより、特許出願における特許請求の範囲や明細書、および拒絶理由通知書等の翻訳精度が高まるということが十分期待できるものと考えます。

このような状況の中、「知的財産推進計画2006」及び「知的財産推進計画2007」の「外国への出願に当たり必要となる特許出願明細書の翻訳作業の際に、誤訳の発生が問題となっていることにかんがみ、一文を短くする、主語述語の対応関係をはっきりさせる、曖昧・難解な用語を避ける等、説明会、解説書等を通じて、誤訳を避けるための明細書の用語や文章の平易化明瞭化を徹底する。」という決定を受け、昨年度、望ましい明細書のあり方について調査を行いました。その調査結果によれば、特許請求の範囲については日本の権利取得を目的とした日本語特許請求の範囲が必ずしも翻訳に適したものと限らず、それゆえ、欧米に出願する場合の特許請求の範囲を考慮した翻訳用日本語特許請求の範囲をあらかじめ作成することが好ましいとの報告がなされております。また、権利範囲を定める特許請求の範囲に適用される産業用日本語と明細書中の技術的事項の記載に適用されるそれとは、それぞれが求められる特性を考慮したアプローチが必要であるのではと考えています。いずれにしても、このような場面においても産業用日本語が活用できる可能性が大いにあるのではないのでしょうか。

進行 横井 : 8月20日 (水) 投稿

特許行政・特許情報における課題とそれへの取組みの中に、産業日本語を的確に位置付けていただきました。産業日本語の役割が明確になり、実体のあるものになってきたと思われます。さらに、特許業務の現状を専門外の者にも簡潔に把握できるようにまとめていただきましたことは、大変、貴重なことであります。

辻井 潤一

東京大学情報理工学系研究科教授／国際言語処理学会（ACL）前会長（2006年）／アジア太平洋機械翻訳協会前会長／アジア言語処理学会連合（AFNLP）会長
機械翻訳、言語処理の研究に従事、現在、東京大学と英国・マンチェスター大学教授、同テキストマイニングセンター・センター長を兼任する



特許文書の翻訳・検索と産業日本語

辻井 : 8月20日（水）投稿

現在、英国に滞在中で産業日本語の報告書が手元にない状態でコメントするので、的外れもあると思いますが、お許しください。

産業日本語という、定義が明確で論理構成が明確な日本語を提案するという基本的な方向には賛成です。また、特許の審査状況や国際的な協力に関する南特許技監の問題点の指摘を読みますと、特許文書の特徴を勘案した、特許文書を志向した取り組みの必要性かと思えます。つまり、**産業日本語という一般的なガイドラインと、そのガイドラインと整合的で、かつ、特許文書を志向した、より詳細な規定（ガイドライン）が必要だ**と思えます。

英語圏でのSimplified Englishなどの取り組みが、今どのような状況になっているのか調べていませんが、以前、機械翻訳との関連で調べた時には、目的に応じて、かなり性質の違ったガイドラインが必要と感じました。たとえば、英語を母国語としないか、あるいは、教育の程度が低い人たちのためのマニュアルでは、できるだけ語彙の数を絞って、限定された単語を使うほうが良いとされます。たとえば、Simplified Englishは、このような観点から使用語彙を絞って、意味の範囲が広いput, take, get, haveといった単語と小辞（up, off, down, etc）とを組み合わせることで、動詞の数を減らす方向に向かいます。ただ、この方向は、機械翻訳や文書の検索という観点からは、あまり望ましいことではありません。このように、一般的に有効なガイドライン（不必要に文が長くなることを避ける、など）のほかに、用途によって異なるガイドラインを設ける必要があると思えます。

また、検索、翻訳をあまり強く意識すると、日本語としては不自然なガイドラインになる可能性も高く、それ

だけでは、日本語としてはかえって読みにくくなることもあるので、適切な補助ツールを用意することが不可欠でしょう。たとえば、文をなるべく短く、かつ、自然な日本語では頻繁に見られる省略を避けるというガイドラインは、計算機処理にとってはよいガイドラインですが、それを読む人間の立場からは、直観的な理解が得にくい、まわりくどい文書になる可能性があります。計算機処理で、あいまいなく補える省略などはゆるして、困難が生じるところのみを明示的に書かせるためのソフトウェアツールを用意する必要があります。現在のGrammar Checkerを一步すすめたようなソフトウェアです。

特許文書の共有に特化したツールとしては、南特許技監が指摘されているような、請求に関する審査官の判断の部分（先行特許との関連など）に関しては、それを明示するメタデータをまず設定し、その種の情報が日本語でどのように表現されるかを分析することで、審査官にそれほど意識させずにメタデータ化することが可能ではないかと思えます。このように、**実際の個々の内容に関する部分（一般的な産業日本語のガイドライン）と、特許固有のメタデータを抽出するためのガイドラインとを区別**することは必要でしょう。

実際の開発とか社会への普及のためには、問題となるような比較的たくさんの文書をガイドラインにそって書き直し、その書き直しが翻訳とか、検索に本当に有効であること、また、書き直した文書が人間の理解も助けることを早い段階で示すことが、不可欠かと思えます。それと、このような観点からは、計算機処理に理解のある国語学者などとの協力も考える必要があると思えます。





IPDLなどの海外向けサービスでは、既に機械翻訳を利用した翻訳が行われているわけですが、南特許技監がご指摘の通り、翻訳精度の点でまだまだ改善の余地はあるものの、技術内容の概略を把握する上で役に立っていると評価していただいているのは、機械翻訳の開発者としてうれしい限りです。今後も引き続き翻訳精度の向上に努めて参りますが、人間の翻訳者に匹敵するような高い精度の翻訳が今後数年のうちに実現できるような技術的ブレイクスルーが得られているわけではありません。コンピュータが人間のように言語を理解した上で翻訳が行えるようになるには、まだまだ時間がかかると考

進行 横井 : 8月20日 (水) 投稿

英国からのコメント、ありがとうございました。言語処理・文書処理の専門の立場からのコメントは貴重です。しかし、手元に報告書の無い状態でコメントをお願いすることになりましたのは、少々残念との感があります。産業日本語は、英語におけるSimplified EnglishやControlled Englishの後追いをするものにはならないではと思われま。しかし、特許文書の特性を分析し、その特性をうまく利用した仕組を設けるべきとのご指摘は、心掛けるべき重要な観点です。

産業日本語に対する期待とプラットフォームの開発

木下 : 8月21日 (木) 投稿

東芝 研究開発センターの木下と申します。機械翻訳技術を研究開発している研究者の視点と、実際に企業の技術者として特許を書く者の視点から、産業日本語に対する期待やプラットフォームの開発について考えていることをお話したいと思います。

日本語が英語のように世界の公用語として通用しない以上、日本の技術情報を世界に発信するためには、英語に翻訳せざるをえません。しかし、あらゆる技術文書をすべて英語に翻訳するには大変なコストがかかり、現実には不可能です。特許も例外ではなく、外国出願するものでない限り、出願人の手で英訳することはほとんどありません。ましてや、年間40万件にのぼる特許明細書を、公的機関がすべて人手で翻訳して公開することなどできません。そこで期待されるのが、機械翻訳の活用です。

ています。

さてここで自分でも特許を書いた経験から考えてみますと、特許明細書とは、改めて言うまでもなく発明の内容を同業の技術者が理解できる程度に開示しその権利を主張するものですから、本来明晰に書かれて然るべきものだと思います。しかし現実には権利範囲が不用意に狭められないようにとの想いから、抽象的な用語や表現を使ってしまうことが少なくありません。そしてそれ以上に、清水理事長がご指摘の通り、非常に長い文が多いのが特徴です。機械翻訳にとっては、一般に、原文の長さが長くなると構文的な曖昧性が急増し、正しい解釈を得ることが困難になります。その意味では、日本の明細書でいきなり先頭に出てくる請求範囲は、人間が一読しただけではほとんど理解できないのと同様、通常の機械翻訳処理では、正しい解釈を得るのは困難です。ただし請求範囲の表現には、その構成順序に規則性を見出すこともできます。いわゆる構成要素列挙形式やジェブソン形式です。これらの形式で書かれた請求範囲は、多くの場合、表現上の特徴を利用して機械的に構成要素や処理手順を独立文として分割することができます。当社の機械翻訳ソフト「The翻訳」でも、このような特徴を利用して請求の範囲を翻訳処理の前に自動で書き換える特許前処理機能を提供しています。

しかし請求範囲にこのような規則性があるとは言え、人間が書く文である以上、記述形式にはバラつきがあります。自動処理のための知識も随時改良していますが、実際の特許文書を処理してみると、分割すべき箇所でも分割されない場合や、本来分割すべきでない箇所でも分割してしまう場合も多々あり、まだまだ改善の余地があると感じています。

さらに実施例に関する説明においても構成要素や処理手順を列挙したような非常に長い文がしばしば見受けられますが、これらの長文は、請求範囲のような規則性は

木下 聡

株式会社東芝 研究開発センター
知識メディアラボラトリー室長
1985年東京工業大学大学院
理工学研究科修士課程修了。
同年(株)東芝入社。以来、
対話システム、機械翻訳システム
等の自然言語処理技術や知識表現
の研究に従事。人工知能学会、
言語処理学会、情報処理学会、
各会員。



ほとんどなく、機械的に分割することはほとんどできません。人間であれば、構成図を参照しながら装置の構成を理解したり、流れ図を参照しながら処理の流れを理解することもできますが、現在の言語処理技術ではこういった人間が行っているようなレベルの理解に基づいた処理は行っていませんので、長文が多い特許文書は機械翻訳にとってハードルの高い対象と言っても間違いありません。また人間は理解できたとしても時間が経つと忘れてしまいます。自分が書いた明細書であっても弁理士さんと相談する頃には、一読しただけでは分からない場合も少なくありません。

このようにコンピュータによる機械翻訳だけでなく人間の知識処理の支援までも目指しているのが、この産業日本語なのかなと考えています。機械翻訳を使う場合、ほんの少し人間が手助けすれば翻訳結果が良くなる場合が多々あります。特に長文を翻訳するような場合には、長い句や節のまとまりが1箇所分かるだけで、不適切な分割を避けられることも多く、結果として訳文の精度を高めることができます。しかし、コンピュータ処理に向けた特別な日本語を人間が無理して書くようでは意味がありません。昨年度のフィージビリティスタディでは、機械翻訳で正しい英訳が得られない特許文に対して翻訳精度が向上するように日本語を書き換える実験を行いました。具体的な書き換え手法は、長文を短文化する、省略された要素を補完する、曖昧な語を言い換える、冗長な表現を削除する、などです。その結果このような書き換えは、多くの場合、コンピュータの理解を助けるだけでなく、人間にとっての理解容易性をも向上させることが分かりました。この結果から、**産業日本語は、人間、コンピュータ双方の歩み寄り**

を促す仕組みとの思いを強くした次第です。

さて今年度のフィージビリティスタディで、私たちは産業日本語での記述を支援するオーサリングシステムの実験ソフトを開発します。このシステムの目的は、普通にかかれた日本語を、意味の等しい産業日本語に言い換えることを支援するものです。全ての処理が自動的に行われるのではなく、人間と計算機とのインタラクションを通して言い換えを行います。先にお話した、請求範囲や長文の分割をサポートする情報は、まさにこのインタラクションによって与えることができるわけです。

オーサリングシステムのおもな構成要素は、日本語解析エンジン、日本語言い換えエンジン、日本語生成エンジンです。日本語解析エンジンは、日英機械翻訳システムのモジュールを活用します。日本語生成エンジンは、英日機械翻訳システムのモジュールを活用します。日本語言い換えエンジンは、辞書と規則ベースとで実現します。どういう言い換えが必要で有効かは、オーサリングシステムの出力である産業日本語の用途に依存します。

また、このオーサリングシステムに日英機械翻訳システムを組み込むことにより、現在より高精度の日英機械翻訳システムが実現できます。**ユーザは入力する日本語文に対して、オーサリングシステムと対話する過程で、さまざまな曖昧性を軽減して高精度の訳文を得ることができると期待**されます。例えば日本語の解析過程では、係り受けの曖昧性の軽減や格助詞の意味選択ができます。英語への変換過程では訳語や構造の候補から正しいものを選択することができます。さらに英語の生成過程では、日本語原文では示されていないが英語訳文には必要な、名詞の単数・複数の区別、定冠詞・不定冠詞、無冠詞などを選択できるようにする予定です。このように、ユーザとシステムが産業日本語を介したインタラクションを行うことで、精度の高い英訳文が作成できると考え





ています。またこの実験を通して、日英機械翻訳向き産業日本語の仕様を策定する予定です。

これまでの私の話では応用として機械翻訳を取り上げましたが、産業日本語が資するのは、単に機械翻訳の精度向上だけではありません。情報検索や文書分類など意味処理を前提とする多くのアプリケーションの精度向上にも同様に貢献していくものと期待しています。

進行 横井 : 8月21日 (木) 投稿

機械翻訳システムの開発・普及に尽力されておられる立場から、産業日本語を適切に位置づけていただきました。また、産業日本語のフィービリティスタディの実績を踏まえた上でのご意見は、大変説得力に富むものです。産業日本語が新たな日本語ビジネス・知識ビジネスへの足がかりとなることが予感されます。

サービスの生産性を絶え間なく向上させるための社会インフラ

橋田 : 8月23日 (土) 投稿

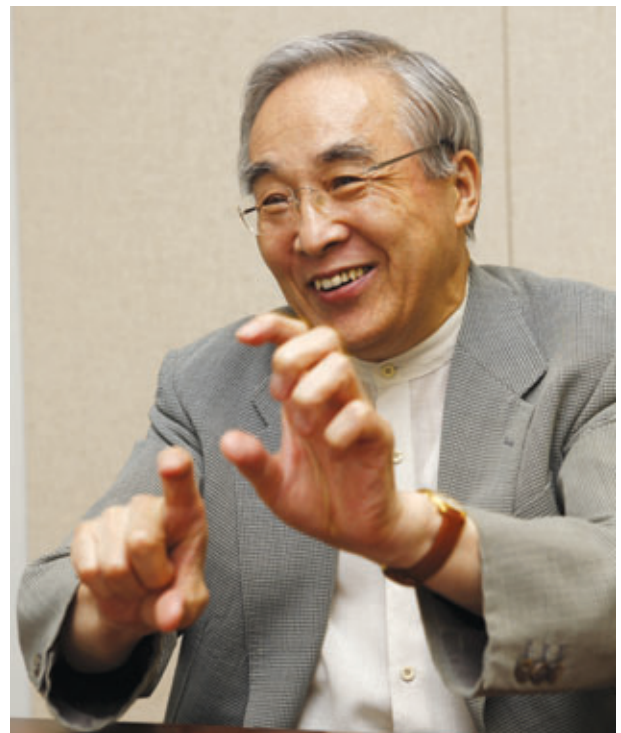
サービス工学の観点から産業日本語を位置付けてみたいと思います。

サービスとは価値を生む社会的（つまり人間同士の）相互作用です。社会的相互作用はモノの受け渡し等の物

理的作用と情報の伝達とを含みますが、そのいずれについても情報技術によって支援または代替できる割合が増しつつあります。社会は人間と情報システムとの共同作業によって営まれていると言えるでしょう。人間は創造的・知的な作業に優れています。一方、人間が得意でない大量で単調な作業をコンピュータは高速・正確にこなすことができます。したがって、人間だけでもコンピュータだけでもできないサービスが両者の共同作業によって実現できるわけです。

この共同作業が成立するには、人間とコンピュータとが意味を共有する必要があります。つまり、人間が理解・操作できる意味とコンピュータが処理できる意味との間に共通部分があり、それを通じて人間の意図がコンピュータに伝わり、人間がコンピュータの出力を理解できなければなりません。

しかし実際には、人間が扱う意味とコンピュータが処理する意味との間の乖離（セマンティックギャップ; semantic gap）により、人間とコンピュータの間で意味が十分に共有されず、両者の共同作業がうまく行かないことがしばしばあります。例えば、Google等の検索エンジンにおいて関係のない情報が山のように出てきたり知りたい情報が現われなかったりするの、人間が何を知りたいのかがコンピュータに正しく伝わっていないからです。また、パソコンのOSをアップデートしたら無線LANに繋がらなくなって1日を棒に振ったなどということが応々にしてありますが、これはソフトウェアの仕組が人間にとってわかりにくいからです。ソフトウェアは人間の生活世界とは無関係にほぼ任意の内容を表現でき



橋田 浩一

独立行政法人産業技術総合研究所
サービス工学研究センター次長
1986年東京大学大学院理学系研究
科博士課程修了。理学博士。1986
年電子技術総合研究所入所。1988
～1992年（財）新世代コンピュ
ータ技術開発機構に出向。2008年4
月より産業技術総合研究所サービス
工学研究センター次長。セマンティ
ックコンピューティングおよびその
応用としての文脈依存型情報サービ
ス、知の社会的共創等の研究開発に
従事。
特許版産業日本語委員会／技術用日
本語プラットフォーム委員会委員長。



るので、人間にとっていくらでも難解になり得るのです。

人間とコンピュータが共有する意味に基づいて情報システムを設計・運用すれば、セマンティックギャップのかなりの部分を解消することができると考えられます。つまり、一般の利用者が理解できるような仕方では情報システムを設計し実装するという方法論です。これをセマンティックコンピューティング (semantic computing) と呼びます。セマンティックコンピューティングとは、人間の生活世界の「意味」を情報システムによる計算処理の直接的な対象にするということです。

産業日本語はCDL (Concept Description Language) という意味記述の枠組に基づいて構造化された日本語です。この構造のゆえに、産業日本語のコンテンツは人間にわかりやすくまたコンピュータによる意味処理の直接の対象にもなります。したがって産業日本語は、セマンティックコンピューティングのひとつの本格的な具現化です。

サービスとは本来は人間同士の相互作用であり、技術や制度によって増幅される場合も創造的な問題解決等はほぼ人間の役割です。したがって、セマンティックコンピューティングの基本的な目的は、人間同士の共同行為 (joint action) を支援することにあります。

共同行為とは、その内容や目的（通常は内容や目的の詳細ではなく概要）に関する参加者の合意（共有信念）の下でなされる行為です。当然ながら、この合意の品質が高いほど共同行為の成功率や品質が高く、合意がなければ共同行為は成り立ちません。サービスも基本的には受容者や提供者を含む人々の中での共同行為です。したがって、サービスの受容者と提供者を含むさまざまなステークホルダの間でのサービスの内容や目的に関する合意は、サービスの生産性におけるきわめて本質的な要因です。

サービスの受容者と提供者との間での商品やサービスの内容に関する合意の重要性は指摘するまでもないでしょう。そもそもサービスの内容が受容者の期待にそぐわなければ意味がありません。また、たとえば治療の目的を患者と医療者が共有することによって治療効果が高まるなど、サービスの目的に関する合意によって受容者と提供者との共同行為の品質が高まり、付加価値が向上します。一方、特に医療等の複雑なサービスにおいては、そのプロセスが関係者の間で明示的に共有されていないことがしばしば生産性の向上を阻害しています。

セマンティックコンピューティングは、人間とコンピュータとの意味の共有を介して、サービスのさまざまなステークホルダの間でのサービスに関する合意（共有信念）の形成を支援します。つまり、**特許明細書や契約書やカルテや規程やマニュアルやサービスプロセスのモデルを意味構造を明示して作成することにより、それを人間にわかりやすく提示したり整合性を自動検証したり翻訳したりすることができ、それが人間の間での正確な理解の共有を促進し、サービスを含む共同行為の生産性を高めるわけ**です。例えば、特許の明細書や拒絶理由書が産業日本語で書かれていれば、それらの意味を出願者や弁理士や審査官等のステークホルダが高い品質で共有することができ、特許の出願や利用に係るさまざまな共同作業の効率と品質を高めることができるでしょう。

サービスは、日本を含む先進諸国において労働人口に関してもGDPに関しても70%に達し、産業のみならず社会全体におけるその重要性が増大しつつあります。産業





日本語は、共同行為としてのサービスの生産性を絶え間なく向上させるための社会インフラとして重要な役割を果たすものと期待されます。

進行 横井 : 8月24日(日) 投稿

サービス工学、そして、セマンティックコンピューティングという最新動向の中に産業日本語を位置付けていただきました。製造業自身も大きくサービス化し、日本の産業全体にとってサービスの生産性向上が必須の課題となるこれから、産業日本語の役割の大きさを予感させてくれるご意見をありがとうございました。

産業競争力強化に繋がる特許情報と 学術情報の間のシームレスと 言語間のシームレスを！

澤井 : 8月25日(月) 投稿

まずは、平成24年度以降での運用フェーズを目指すこの産業日本語プラットフォーム開発計画という野心的なチャレンジに敬意を表します。

しかし、敬意を表したからといってネット座談会に参画などしてはいけないのだということに気付きました。「産業日本語」という聞き慣れない用語を前に、何から書き出せばよいのだろうと迷っています。こんな頭の中を明晰にしてくれるのが「産業日本語」であれば助かるのでしょうか・・・。

冗談はさておき、守屋所長のお話の中に、Japioでの事業立上げ時期に「産業日本語」「明晰日本語」「技術用日本語」という三つの用語が入り乱れていたのを7月18日から「産業日本語」に統一した、との下りがあるのを

P R O F I L E

澤井 敬史

NTTアドバンステクノロジー株式会社 取締役/知的財産ビジネスセンタ所長

現職はNTT-AT知的財産ビジネスセンタ所長。社外では、日本知的財産協会理事長、経済産業省産業構造審議会臨時委員等を歴任し、現在、経団連産業技術委員会知的財産部会部会長代行、東京理科大学知的財産専門職大学院教授(兼務)。



見つけ、その迷いが少し解けたような気がしました。

約10年前に日本知的財産協会で、コスト削減の観点から当時の機械翻訳を特許実務に導入できるか否かを検討したことがあります。そして、その時に感じたことをJapioの特集に寄稿したことがありますので、最初にその一部を引用します。

----- 機械翻訳WGを通じて筆者が素朴に感じた驚きは、機械翻訳ソフトの技術者と日本語ワープロソフトの技術者との間に交流が殆どないということであった。多分、最近では日本語ワープロソフトもかなり進化しており文法の間違いまで指摘するようになっていたので、両者の連携が図られた機械翻訳のソフト開発も進んでいるものと思われるが、技術的には異なるとされる両分野の緊密な協力関係から革新的な機械翻訳ソフトが近い将来に出現することを期待している。-----

今回の「産業日本語」への取組は、まさにこのような両分野の協力関係だけではなく、サービス提供まで視野に入れたより広いシステムを構築しようとしているものだとも個人的には感じています。

ところで、私への要請は、企業の知的財産戦略の視点から「産業日本語」を論じて欲しいというものでした。しかし、正直言ってこの開発計画が壮大であるが故に、それを論じるのはとても難しい気がしています。そこで、その感じている難しさの根っこにある疑問と、それに対する自分なりの感想を幾つか書きますので、今後の議論の材料にして頂ければと思います。

「産業日本語」とはどこまでを意識した言葉なのか？

「産業日本語」の対象文書はどこまでを視野に入れているのか？

「産業日本語」を検討する最終ゴールはどこなのか？

多分、この開発計画がJapioの取り組みとして行われるものである以上、「特許情報」を中心に据えているのだろうと思いますが、特許を超えて産業という言葉まで広げ



ているのは、それなりに意味があることなのでしょう。

しかし、あえて誤解を恐れずにいえば、「産業日本語」と「特許情報」は、知的財産と特許の関係に似て、トリッキーな側面を孕んでいると感じています。一般に、特許の分野で仕事をしてきた人達は、知的財産と言いながら特許の実務を思い浮かべて話をします。一方では、知的財産と言いながら著作権の実務を思い浮かべながら話をする人もいます。これと似ていて、「産業日本語」と言っても、純粋に日本語の構造解析を念頭において話をする人もいるでしょうし、機械翻訳のことを念頭に置く人もいるでしょう。また、同じ技術に関する日本語でありながら、学术论文を思い浮かべる人もいるでしょう。「特許情報」の根幹である特許明細書を思い浮かべる人もいるでしょう。そして、特許明細書であっても、実は分野によってはかなり書きっぷりが違います。これらのどこに焦点をあてるかが、最初の課題です。

おそらく、このあたりの課題については、まずは「特許情報」に重点を置き、開発計画が進められるものと推察しています。また、分野別の特徴への対応に関しては、Japio 2007 Year Book の横井顧問の発言などを讀ませて頂くと、ひとつの日本語仕様を定めるのではなく、それぞれの分野でそれぞれの目的に応じた仕様が定められる方向で、検討が進められるのであらうと感じています。

次に、「産業日本語」の対象文書についてです。対象文書としては、特許情報の最たるものとしての特許明細書、いわゆるノウハウなどの技術情報を文書化した技術文書、その他には研究論文等の学術文書、などが考えられます。論理の組み立て方や用語の使い方などにおいて相当な違いのあるこれらの文書を、全く同列で同時に検

討していくのか、それとも優先順位をつけて検討していくのかも、大きな課題になります。

先にも書いたように、たぶん特許明細書を中心に検討が進められるものと思いますが、できることなら非特許文献の最右翼である学术论文も同時並行的に検討をして欲しいものだと考えています。その心は、清水理事長が述べられているところと全く一致しています。言うまでもなく、イノベーションのブレイクスルーは研究分野から起こり、それは学术论文という形で公表されます。その後、企業などの努力により製品・サービスという形でマーケットに提供される過程で数々の技術が特許明細書になります。その基本フローは変わりませんが、これからグローバルな競争がますます熾烈になる中では、学から産へのバトンタッチがスムーズに行われるようにすることが、国際的な産業競争力を高める上で必須でしょう。そのためには、**学术论文と特許明細書が同じような視点に立って理解しあえる土壌を作ることが必要**です。それに向けて「産業日本語」は、既存の特許明細書の悪い類型にとらわれることなく、清水理事長のおっしゃるように**“一貫性があり、平易で簡素・正確な記述”**を特許明細書の世界にも導入する観点から検討されることを期待しています。

最後の課題は、「産業日本語」を検討する最終ゴールは、どこにあるのかという点です。これは、守屋所長のお話にあるように、産業日本語を我が国の産業技術文書に広く導入し、我が国の産業技術情報の情報力を飛躍させ、我が国の知的生産性の更なる向上と経済活性化を期待すること、を狙っているのだと思われます。

そこで、情報という観点から考えてみると、企業には情報に関して二つの立場が常に存在します。ひとつは産業技術情報（特許情報と置き換えた方が正しいかもしれませんが）を創り出す（特許出願する）立場、もう一つはすでに公表されている産業技術情報（この場合には、特許情報だけでなく学術情報を含む広い意味の方が適切でしょう）を調査・分析して利用する立場です。

情報を創り出す立場の企業に、「産業日本語」を使ってもらうには、かなりの工夫がいるのではないのでしょうか？ それを**採用する際のインセンティブを、経済性、利便性、必要性の三つの視点から考えてみる**ことが肝要でしょう。

経済性は、まさにそれを使えば従来のものよりも安くできることが指標になります。ただし、安かろう悪かろう（使いものにならない）ではいけません。

利便性は、それを使うと楽である又は効率が良かったことが指標になります。これは、経済性とのセットで評価されることになるでしょう。

必要性は、それを使うことが社会的なルールとなっているかどうかということが指標になります。ただ、これには注意が必要で、コスト高をもたらずルールでは

駄目で、ルールが技術の現状に歩み寄りながらも実務に支障をきたさないことがポイントになります。

このあたりについては、別の機会があれば例を挙げてご説明したいと思います。

また、産業技術情報を利用する立場から言うと、守屋所長が開発計画の三番目に挙げているアプリケーションシステムにおいて、知的で高度なソフトを開発してほしいということに尽きます。

その際には、特許情報だけでなく学術論文を含めた科学技術情報の総体が対象となるものであることが望まれます。清水理事長がおっしゃっているように、両者の間をシームレスに検索できることが大切です。そのためには、研究者が論文を書くときに特許明細書を参照する習慣、逆に技術者が特許明細書を書くときに論文を参照するような習慣を醸成するようなことも、運動論として大事だとかねがね考えています。

そして、シームレスは実はもう一つあって、清水理事長も指摘されている言語に関するものです。科学技術情報を巡ってグローバルな競争が起こっている現代ですから、いうまでもなく、この開発計画の対象エリアには日本語情報だけでなく外国語情報も入っていることでしょう。従って、機械翻訳を例にとれば、日本語を中心に据えた機械翻訳を深化させることで、**日韓間や日中間の機械翻訳を行なえる環境を日英間と同等以上のものにまで高め、日本語を中心にした機械翻訳のアジアワールドを速やかに実現するような戦略的な視点**も忘れてはならない重要なことだろうと思います。

「産業日本語」が、特許情報と学術情報の間のシームレスと、言語間のシームレスとを実現したときに、日本の産業がグローバル世界でより強い競争力を獲得していることを大いに期待したいと思います。



進行 横井 : 8月25日 (月) 投稿

企業の特許業務の現場におられる立場から、計画に対する心構えを含め産業日本語を位置付けていただきました。現場の方々のニーズを適切に汲み取ること、実務の現場の協力のもとに進めなければならないこと、これらのご指摘は、産業日本語の開発・普及に際して常に心掛けねばならないことであると思っております。

日本の情報システムを日本の特許情報システムからアプローチ

進行 横井 : 9月3日 (水) 投稿

守屋所長の『産業日本語プラットフォーム開発計画』を巡る形で、皆様のご意見を一巡させていただきました。『産業日本語』を位置付ける重要な論点は、すべてご指摘頂いたと思われまます。また、『開発計画』を具体化するに当たって留意すべき要件も適切に論じていただきました。

次に、皆様に二巡目のご意見を伺います。一巡目で頂いたご意見を加味しながら論点、あるいは、要点をまとめてみます。このまとめを参考にしながら、再度のご意見を申し上げます。一巡目のご意見を発展させる形でも結構ですし、全く新しい観点からのご意見でも結構です。そして、皆様のご意見を伺った最後に、守屋所長に『開発計画』を具体化していくための考え方をまとめていただき、本ネット座談会の締めとさせていただきます。それでは、参考にしていただくための論点・要点を以下にまとめます。

① 産業日本語は多様である

産業日本語は、ひとつではなく、それぞれの用途ごとにそれぞれに日本語仕様が必要です。そして、それらの日本語仕様は、共通となる部分とそれぞれに個別となる部分に分かれることとなります。ただし、そのような産業日本語全体を最初から開発するというわけにはいきません。そこで、特許情報をターゲットに、今までの蓄積を素直に生かせ、そして、効用がはっきりと分かるいくつかの産業日本語事例からアプローチすることになります。現在、Japioでは、**日英機械翻訳向き産業日本語(翻訳TJ)**と**特許ライティング向き産業日本語(特許TJ)**に取り組んでいます。なお、産業日本語の英語名は、**Technical Japanese (TJ)** です。

② 日英機械翻訳向き産業日本語(翻訳TJ)

翻訳TJは、現在の機械翻訳システムが正確に情報を伝える英文に翻訳できる原文和文を書くための日本語仕様です。かつて、米国では、不完全な機械翻訳システムに対応するため簡易化英語や制限英語で原文英文を書かせ

るというアプローチがとられました。しかし、この試みは成功しませんでした。我々の翻訳TJは、米国でのアプローチとは逆です。我々は、すでに、相当の機能を持つ機械翻訳システムを利用することができるのです。その機械翻訳システムの前編集機能を体系的に手順だったものにする仕組として、翻訳TJとそのオーサリング機能を用意するという考え方です。この**オーサリング機能は、非明晰な日本を明晰な日本語に翻訳するインタラクティブな機械翻訳機能として実現されることになるでしょう。**

③ 特許ライティング向き産業日本語（特許TJ）

特許TJは、**特許明細書等を明晰な文書として書くための日本語仕様**です。望ましい明細書のあり方や書き方について、十年周期ぐらいで調査検討が繰り返されてきました。しかし、得られた結果には、あまり大きな変化は見られないようです。大きく前進させるためには、発想の仕方を変える必要があります。まず、特許文書の問題点の分析には、特許・知財の専門家と言語学・言語処理の専門家が協力して行う体制が必要です。その結果を、まず、出来るだけ詳細に渡り手順だった特許文書ライティングマニュアルにまとめることが必要です。そして、次に、ライティングマニュアルに基づいて、オーサリングシステムや共通辞書などのライティング支援環境の整備へと進むことになります。この時、重要なことは、標準化とか規則化などということを最初に持ち出すことはしないようにすることでしょう。**ライティングマニュアルやライティング支援環境は、知財の現場の多様な意見やニーズを柔軟に反映しながら斬新的に進化していく仕組とすべきでしょう。**

翻訳TJと特許TJを一体化したTJを設けると、特許文書を正確に情報を伝達できる英文に翻訳できる産業日本語が得られることになります。特許TJは、当然、特許文書に特化された産業日本語ですが、翻訳TJは、出来るだけ広範囲の技術文書に適用できる産業日本語であることが望まれます。

④ 人にも機械にも明晰である

産業日本語は、人にも機械にも明晰であることが重要な要件です。例えば、**翻訳TJは、一般の技術文書として人に違和感のない日本語でなければなりません。また、翻訳TJは、機械翻訳システムという機械が正確な英語に自動翻訳できる日本語でなければなりません。**明らかに、このような**人に対する明晰性と機械に対する明晰性をひとつの日本語として実現することは無理**というものです。そこで、**人に対する日本語と機械のための日本語を設け、この2つの日本語を連携処理できる仕組**を設けることになります。機械のための日本語には、形態素解析、構文解析、係り受け解析などの解析処理の結果や翻訳処理に必要な情報が付加されます。この機械のための付加情報を記述し、連携処理の仕組を提供してくれるのがCDL (Concept Description Language) です。とりあえ

ずは、CDLは、XMLを発展させたものであると思ってください。

⑤ 特許文書も多様である

特許関連文書は、申請側の文書と審査側の文書に分かれます。申請側の文書は、大きく4種類の書類、「特許請求の範囲」「明細書」「図面」「要約書」から成り立っています。審査側の文書を代表するが「拒絶理由通知書」です。これらの文書や書類は、それぞれ役割が異なりますから、日本語の用法も異なることになります。「特許請求の範囲」は**権利文書**ですから、**強い権利を広い範囲で主張する論理的に明晰な日本語**が必要になります。「明細書」は**技術文書**ですから、**通常技術系学術論文を書くための日本語に出来るだけ近い日本語**が必要になります。「要約書」は**レファレンス文書**ですから、**要約記述のための簡潔で平易な日本語**が必要になります。「拒絶理由通知書」は**権利判断文書**ですから、**権利に対するメタな記述を行うための論理的に明晰な日本語**が必要になります。そして、特許の技術分野ごとに、特許関連文書に特色が生まれます。用語が異なってくるのは当然ですが、構文や文章構成に関して日本語の用法がどの程度異なってくるのか、これについては正確な分析が必要です。いずれにしても、どの文書や書類に対する日本語についても、文書検索や機械翻訳に対する高い適合性が要求されます。

⑥ 日本の情報システム

特許情報をより高価値化していくため、特許の申請・審査をより精度の高いより効率の良いものにしていくため、申請・審査のグローバル化に対応していくため、特許関連文書の日本語を大きく見直そう、まずは、それが手始めです。そのためには、色々な技術開発も必要です。



しかし、重要なのは、**社会的なシステムとして、あるいは、社会的なサービスとして、あるいは、場合によっては、社会的な制度として、どのような手順で実現し、どのように定着させていくのか**、このあたりに対するコンセンサス作りであると思われます。日本の情報システムの根幹となるのが日本語です。しかし、世界の情報システムの根幹は、色々曲折はあるにしろ、英語に収斂していくでしょう。日本の情報システムと世界の情報システムの連携のためにどのような仕組みを作り上げ、日本の情報システムをどのように発展させていくのか、そこに、産業日本語の最大の位置付けがあります。そして、日本の情報システムを日本の特許情報システムからアプローチしてみようというのが、Japioの提案です。

言語の壁を越えた 科学技術情報システムに

清水 : 9月8日 (月) 投稿

門外漢の私にとって、このネット座談会における諸氏の見解は新鮮で、この分野が「旬」であることを感じさせられました。最後の「まとめ」も要領よく意見が集約され、今後の活動の指針になることでしょう。

敢えて言うならば、「6:日本の情報システム」の中に記載されている一文“世界の情報システムの根幹は、色々曲折はあるにしろ、英語に収斂していくでしょう”が気になります。

私の個人的体験からも、科学技術の研究成果は「英語」で書かない限り、世界の関係者に読んで貰いにくいことは確かです。また、国際会議も「英語」を使用する必要があります。したがって、学生諸君には、日頃から「英語」の論文を読み、「英語」で成果を報告する訓練に多くの努力を払うことを強いてきました。

ただし、「英語」が特に科学技術表現に適した言語というわけでもないのです。科学技術は、「客観的、合理的な説明ができ、いつでもどこでも実証ができること」をベースに展開されていて、どの言語を用いても相互に収斂できる本質を有しているのです。したがって、世界の趨勢と利便性のためだけに「母国語」を放棄し、「英語」に走ることは賛成できません。音楽やある種のスポーツと同様に、科学技術においても創造性を高めるためには、小さい頃から格調の高い科学技術情報に接することが大切です。「英語」の読み書きが出来てからでは遅いのです。

その点、「知的財産情報」は地属性が高く、日本の特許は「日本語」で書くことが原則です。しかも、毎年5万件近い特許が外国から日本特許庁に出願されるそうです。現在は、情報の集積地だけが、「情報システム」の運用のチャンスが与えられます。そこで、特許等の「知的財産情報」と、関連する科学技術情報をシームレスに連結さ



れたシステムが実現できれば、**言語の壁を越えた科学技術情報システム**に一步近づくことになるでしょう。

進行 横井 : 9月8日 (月) 投稿

私のまとめにあります「日本の情報システムの根幹となるのが日本語です。しかし、世界の情報システムの根幹は、色々曲折はあるにしろ、英語に収斂していくでしょう。」というくだりは、明らかに言葉足らずでした。「世界の情報システム」というのは、「世界中の情報システム」という意味ではなく、「世界が共有して使う部分としての情報システム」という意味で、ご意見の趣旨と同じです。しかし、言葉足らずの表現であることは、そのとおりです。適切なお指摘ありがとうございました。

産業日本語の普及のために 使用を推進する活動を

木下 : 9月11日 (木) 投稿

澤井さんのお話の中に、『情報を創り出す側の企業に「産業日本語」を使ってもらうには、経済性、利便性、必要性という三つの視点でインセンティブを考えるべきだ』とのお指摘がありました。

経済性という点では、外国出願の際の翻訳コストがいかに減らせるということが一番分かりやすい評価指標でしょう。機械翻訳の出力である限り最終的には人手によるチェックは必要でしょうが、産業日本語で書かれたテキストとそうでないテキストの翻訳コストの違いは定量的に評価できるので、まずはこの経済性でメリットを訴えられるようにしなくてはならないと思っています。



一方利便性を考える際の課題は、特許情報の作成者が、産業日本語で執筆するメリットを感じられるまでに時間がかかるという点だと思います。知財担当者や特許事務所の方々は明細書がより理解しやすくなるという意味ですぐに利便性を享受できるでしょうが、執筆者にしてみれば作成時に余計な手間が増えただけと思われかねません。公知例調査の時間が短くできるといった執筆者にとってのメリットが必要です。特許提案の際は、通常明細書を執筆するよりも公知例調査に時間がかかる場合が多いですから、執筆の時間が多少長くなっても、それ以上に調査が簡単に済めばトータルでは利便性があると判断できます。しかし出願した特許が公開され、しかもある程度の量が貯まってこないと調査作業の効率化は実感できないでしょうから、一定の年数が経過しないと執筆者がメリットを実感するのは難しいように思います。当然のことながら産業日本語での執筆をサポートするオーサリングシステムをなるべく使いやすいツールとして提供しなくてはいけません。最初のうちには強制ではないにしろかなり強力に使用を推進する活動が必要ではないかと考えています。また以上のことから、特許検索のサービス事業者が、既に出願されている過去の膨大な特許に対してメタ情報として産業日本語化情報をつけた独自の特許データベースを構築し、より高精度な検索を謳い文句にビジネスを展開するといった活動が産業日本語の普及には重要かもしれません。

進行 横井 : 9月11日 (木) 投稿

産業日本語の開発・普及には、環境づくりが必要であるとのこと指摘は、大変、貴重です。産業日本語の効用が実感できる環境づくりには、特許情報環境は、先進的なモデルとなってくれるのではないかと思います。

特許文書ライティングマニュアルやライティング支援環境の整備

澤井 : 9月12日 (金) 投稿

横井先生が、皆さんの意見を①から⑥まで上手にまとめて下さったので、全体の構図が文字通り、より明晰になりました。

今回の開発計画は壮大であり野心的なものであるだけに、⑥にもあるようにコンセンサス作りが大切でしょう。そのためには、**先ずは限定的でも良いですから、実務家が使いたくなるものを早く提供することが一つの手でしょう。**その観点で①から⑥までを一読してみると、初期の段階において、特許情報の中でも特に“明細書”に絞り込んで、検討を進めてみるのが効果的であるという印象を持ちました。

特許公報は、権利文書と技術文書という二面性を持っていると良く言われます。

“特許請求の範囲”は、⑤にあるように権利文書の根幹をなします。特許独立の原則があるので、独占排他的な効果を得ようとするならば、各国ごとに特許権を取得する必要がありますし、現実には他人を排除しようとするならば、その特許権に基づいて各国の裁判所で争わなければなりません。従って、権利行使に深く関わる“特許請求の範囲”を、いまずぐ明晰化の対象とすることには、実務家のアレルギーが高いですし、コンセンサス作りには時間がかかると思われれます。

一方で、“明細書”は技術情報を社会に広く知らしめ



る役割を持っています。そして、③には「特許TJは、出来るだけ広範囲の技術文書に適用できる産業日本語であることが望まれます。」、④には「翻訳TJは、一般の技術文書として人に違和感のない日本語でなければなりません。」、⑤には「“明細書”は技術文書ですから、通常の技術系学術論文を書くための日本語に出来るだけ近い日本語が必要になります。」と書かれていますので、先に述べた絞込みを違和感なく行える共通認識ができています。また、科学技術の世界における「一貫性があり、平易で簡素・正確な記述」の方法論を、技術文書としての性格を持つ“明細書”に移植することは、比較的容易に行えるでしょう（といっても実際は大変なことです。が・・・）。従って、③において述べられている**特許文書ライティングマニュアルやライティング支援環境（オーサリングシステムや共通辞書など）**を、まずは“明細書”に関して整えることにすれば、大いに利用価値が高くコンセンサスが得やすいと思われます。

最後に余談ですが、最近ではインターネットの普及でコピー＆ペーストで文章を作成すること）に頼る文書が増えており、自分で論理を構築し文章を書くことの力が全般的に落ちてきていると感じることがあります。本題とはやや外れるかもしれませんが、今回の開発計画を進めるに際しては、“書く”という基本能力を高めるような方策が並行して講じられると良いなとも感じています。

進行 横井 : 9月13日 (土) 投稿

明細書から段階を踏んでというご提案は、説得力のあるご指摘です。特許文書全体を明晰化するには、まず明細書を一般的な技術文書に近づけることから始めるのが効果的であるとのことご指摘です。いわゆる学術論文も、日本語テクニカルライティングの観点から見れば、多くの課題を抱えています。意外に、明細書の明晰化が一般の学術論文の明晰化にも貢献できるということになるかもしれません。

「特許庁業務・システム最適化計画」における特許庁の検索システム

南 : 9月16日 (火) 投稿

皆様より産業日本語について文書検索の観点からの御意見もありましたので、現在の特許庁の検索システムおよび目下検討が進められている「特許庁業務・システム最適化計画」における特許庁の検索システムについて紹介させて頂きたいと思っております。

これまで特許庁では増大する特許文献を効率的にサー



チするために、Fターム、テキスト検索システム等を開発してきました。Fターム検索では特許文献の技術的内容を予め解析して多観点の検索インデックスを付与しているため関連のある文献を効率的に絞り込むことができます。そして、現在、このFターム検索とテキスト検索と組み合わせることにより従来に比べて先行技術文献調査の大幅な効率化を図ることが可能となっています。

また、上述の最適化計画では特許庁のシステムを特許庁運営基盤システム（平成24年稼働）と特許庁新検索システム（平成26年稼働）の2つに分けて開発していくことになっていますが、後者の特許庁新検索システムでは世界最高レベルの迅速かつ的確な審査を行うために**概念検索技術、審査ナレッジ及び外国文献に対する機械翻訳等の最新の情報技術を採用した検索が行えるよう検討を進めている**ところです。

このうち概念検索について、特許明細書の文章を産業日本語化することによる検索の高度化の可能性はすでに報告書「明晰日本語」において言及されているところですが、明細書作成時だけでなく、**既存の産業日本語で記載されていない大量の特許文献についてもテキストデータが存在し、効率的に産業日本語に変換することが可能となれば概念検索のさらなる高度化が図られる可能性**があると思われま

先に述べました機械翻訳への適用等も含めて産業日本語の大いなる可能性を鑑みると、実用化された産業日本語が広く産業界に受け入れられ、特許法の目的でもある産業の発達に寄与する日が一日も早く訪れる日が来ること、および、それを実現するために重要であるJapioの継続的なリーダーシップの発揮を期待しています。

進行 横井 : 9月16日 (火) 投稿

常に最新の情報を反映させながら統合的に管理・運用されている産業情報の環境・システムとしましては、特許情報の環境・システムは、質・量共に群を抜いたものです。特許庁が検討を進めておられる諸計画に産業日本語が適切に貢献できますならば、特許情報と広く一般の産業技術情報をいろいろな側面で媒介する役割を産業日本語が果たせるのではないかと、そのような期待をこめたご指摘をいただきました。

グラフ表現形式によるオーサリング方式のアプローチ

橋田 : 9月26日 (金) 投稿

コンテンツを作成した後に手間をかけて明晰化するのではなく、初めから明晰な形で作成することによって作成が容易になる可能性があります。特許明細書の明晰化と作成の効率化を同時に達成する可能性について、昨年度の報告書で使った例を再利用して少し考えを進めてみたいと思います。

ある特許明細書の請求項の部分を構造化したものを図示します。請求項は3つあります。第1の請求項は3つの構成要素からなり、第2の請求項は1つの構成要素からなることがわかります。第3の請求項は前提部と2つの構成要素からなりますが、これら2つの構成要素は第1の請求項の2つの構成要素の目的と同じものです。さらに、第1の請求項の第2の構成要素および第3の請求項の前提部はほとんど当たり前の内容なので、これらの請求項の内容は実質的に同じですが、第1の請求項が装置に関するものであり、第3の請求項が方法に関するものである点だけが異なります。

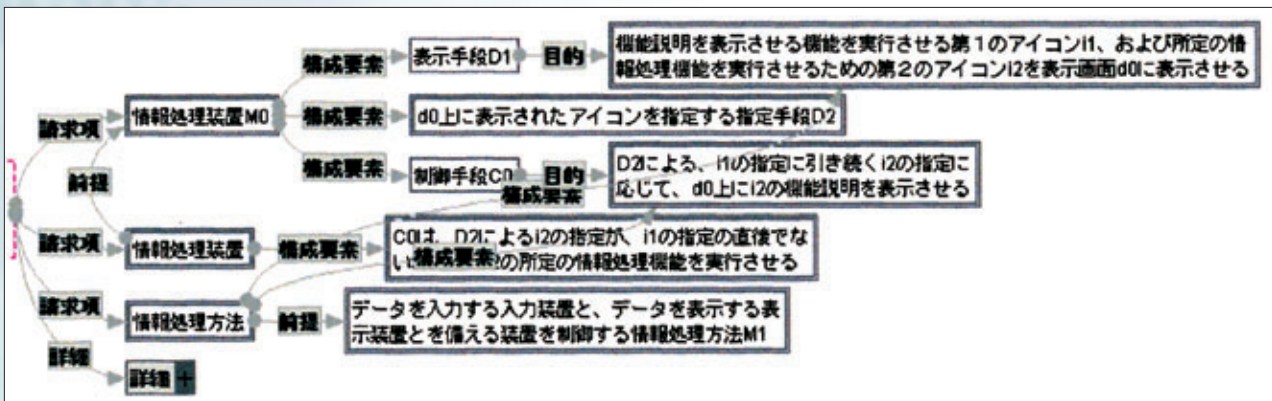
このように、請求項の文書構造のオントロジーに基づいて請求項の内部構造と相互関係がわかりやすく整理でき



るので、それによって特許明細書の作成・編集が容易になり、その品質も向上するものと期待されます。

もちろん、図のようなグラフの形のコンテンツは今のところそのまま特許庁に受理してもらえないので、このようなコンテンツから通常のテキストを生成する必要がありますが、その作業は自動化することができます。たとえば最初の請求項は「機能説明を表示させる機能を実行させる第1のアイコンi1、および所定の情報処理機能を実行させるための第2のアイコンi2を表示画面d0に表示させる表示手段D1と、d0上に表示されたアイコンを指定する指定手段D2と、D2による、i1の指定に引き続くi2の指定に応じて、d0上にi2の機能説明を表示させる制御手段C0を備えた情報処理装置M0」のようになります。

このようなグラフ状の構造化はもう少し細かくできるかも知れません。それはおそらく、ノードが単文でリンクが談話関係であるようなグラフになるでしょうが、特許明細書の場合には談話関係の種類を10程度に限定でき



るのではないかと考えられます。もっと細かいレベルの明晰化はこのような構造化とは独立して扱うことができるでしょう。しかし、このような粒度の粗い構造化が、より細かいレベルの明晰化の助けにもなると考えられます。

進行 横井 : 9月26日 (金) 投稿

請求項のような複雑な構造を持ったテキストに対して、グラフを用いた表現形式は、その複雑な論理構造を明晰化するのに役立ってくれると思われます。さらに、グラフ形式化することによって自然言語の持つ線状性という制約を克服し、それに適したオーサリング方式を工夫するというアプローチは、産業日本語にとっても色々な可能性を予感させるアプローチです。ただ、現場の方々に違和感なく使っていただけるようになるには、まだ、色々な工夫が必要であると思います。

辻井 : 9月29日 (月) 投稿

まとめにあるように、産業日本語が設定する一般的な原則のもとに、個別用途用の、より既定の詳細なものを設けるとするのは賛成です。これとも関連しますが、一巡目のSimplified Englishの議論で言いたかった点は、人間にとっての簡単さと機械システムにとっての簡単さが必ずしも同一ではない、ということでした。明晰性というの、何の目的で、誰を対象とするかで変化するもの、ということです。

目的や処理システムを考えずに明晰性を議論するのは、私も誤りだろうと思います。この点、まとめにあるように、特許の機械翻訳という具体的な問題からアプローチするという方向に基本的に賛成です。ただ、最近の機械翻訳の動向を見ると、統計ベース、例ベース、規則ベースといった異なったパラダイムの翻訳システムが開発されてきていますから、これらすべてのシステムに**共通する明晰性というの、うまく定義できるのか**、というのが気になります。産業日本語の設計方針を立てるときに、議論すべきことであろうか、と思います。

また、清水さんが提起された日本語、英語の問題も、考えておく必要があるでしょう。横井さん、清水さんの指摘のとおり、特に、特許という国家戦略に密接に関連する情報がすべて英語化されることはないでしょうから、常に、多言語の環境下での情報の収集と発信ができるプラットフォームを考えておく必要があると思います。特に、木下さん、南さんの指摘のように、産業日本語プラットフォームが広く使われるためには、その利便性の向上が不可欠、とくに、先行特許、類似特許などの

検索系と有機的につながることが不可欠かと思います。このとき、**日本語を英語へ、という方向の翻訳も重要ですが、既存の外国特許の検索と日本語への翻訳の機能も、本当は議論しておく必要があるか**と思います。

南さんのFタームとテキスト処理の議論は、翻訳とは別の視点から、産業日本語を考える上で重要だと思いました。専門家が設定する制限された用語 (Controlled Vocabulary) でテキスト内容を規定するというのは、やはり検索系にとっては強力ですから、このようなメタデータの記述をもう少し構造化して記述しておく、その際の産業日本語とは何かという問題です。産業日本語でテキストの内容すべてが記述されるというより、その特許を特徴づけるPhraseやClauseが、テキストが産業日本語でかかっていると抽出しやすくなる、といった特徴付けができるとよいと思います。この観点からは、澤井さんの議論の請求範囲、明細書のどちらが対象として有効かは、素人で判断できません。また、内容すべてではなく、構造化された骨組のようなメタデータであれば、橋田さんが示すような図式表示も、わりと使用時の抵抗が少ないのではないかと思います。

進行 横井 : 9月30日 (火) 投稿

お忙しい中、二巡目のご意見をありがとうございました。いずれにしましても、産業日本語は、大テーマです。日本は、外国人日本語教育のシステムを作り上げ普及させるのに20年ほどの時間を要しました。産業日本語も、産業界に広く行き渡る仕組を作り上げるには、やはり、同じぐらいの時間が必要であると思われます。

また、そのためには、言語学や言語処理の専門家の



方々の知恵を結集する仕組みを作り上げることが必須であろうと思っています。

Japioは、産業日本語という大テーマに、特許情報という側面から取り組むことになります。特許情報自身も非常に大きく複雑なものですから、着実なステップを踏むためには、さらに切り込んだ取り組みが必要になります。守屋所長に「産業日本語プラットフォーム開発計画」へどのような糸口から取り組み、確実なステップを踏んでいくのかをご説明頂き、本ネット座談会のまとめとしたいと思います。

おわりに 産業日本語を具体化するための開発構想

守屋 : 9月30日 (火) 投稿

産業日本語に関するご意見やご要望、開発を進めるにあたっての留意点や課題等についての貴重なご指摘をいただき、誠にありがとうございました。皆様のご意見を踏まえつつ、最初にご説明した『開発計画』をJapioが今後どのように具体化していくべきか、現時点での構想を述べさせていただきます。

『開発計画』における開発項目として挙げた3つの基盤技術、すなわち「プラットフォーム仕様」「プラットフォームシステム」「アプリケーションシステム」は、あらゆる産業技術文書に共通して適用可能なものです。ただし、これらはあくまで基盤であり、産業日本語本来の効果を得るためには、個々の産業技術文書の特性に応じ、これをチューンナップしたり、必要な機能の追加やデータの蓄積を行う必要があります。

Japioの専門性は特許情報であり、産業日本語もまずは特許関連文書における普及を大目標として開発を進めるべきと考えておりますが、その際、まず基盤技術を開発し、それから各基盤技術の特許文書に応用していく、という順序では、実用化までに相応の年月が必要となります。それよりはむしろ、部分的・小規模であれ、**開発当初より産業日本語を実際の特許関連文書の作成現場に導入するための開発を行い、その効果を実証しつつ、徐々に範囲を広げていくという形で進めるのが現実的**と思います。つまり、『開発計画』にて最後に挙げました「特許関連文書を対象としたモデル運用サービス」を実際の開発順序としてはいちばん先に進めることを想定しています。基盤技術に関しては、こうして特許関連文書を対象に開発した技術から、他の産業技術文書に転用可能な部分を抽出していけば、効率的にその土台が得られるでしょう。

ただし、仮に特許関連文書を対象とした技術、例えば、

特許明細書用のオーサリングシステムを開発できたとしても、いかにこれを特許業界に普及させていくか、という別の課題が残ります。澤井取締役からもご指摘のあったとおり、経済性、利便性、必要性に係る強いインセンティブがない限り、非常に難しいことは確かです。

インセンティブという観点を別にしても、作成者が不特定多数である特許明細書は、特許関連文書への産業日本語の導入の糸口としてはあまり適さないように思われます。よってJapioでは、『開発計画』の具体化にあたり、まずは、作成者が限定でき、かつインセンティブが存在すると思われるものを対象としていくべきと考えます。

一例を挙げますと、Japioが受託しております公開特許公報英文抄録 (PAJ) の作成工程への導入を前提に、産業日本語のオーサリング技術の開発を検討しております。現在、PAJはほぼ100%人手翻訳にて作成しており、機械翻訳は参考情報として利用しているものの、大きな作業効率化は達成できておりません。その原因として、原文である出願人要約が機械翻訳処理に向いておらず、結果として精度の高い機械翻訳が行えないことが挙げられます。産業日本語のオーサリング技術は、非明瞭な表現や曖昧な表現をアラート (注意喚起) し、インタラクティブな形で修正を促すものです。これをPAJ翻訳の前処理に導入し、出願人要約を明晰化すれば、精度の高い機械翻訳が可能となり、人手翻訳の比重を軽くできるのではないかと考えます。この試みが成功すれば、PAJ作成事業の最大の課題である機械翻訳の利用による効率化が果たされるとともに、特許文書用オーサリングシステムのプロトタイプが得られることとなります。加えて、



産業日本語及びプラットフォーム技術の効果を、特許庁や特許業界に示し、新たなニーズの創出につなげられる可能性があります。

PAJ作成への産業日本語の適用に関連して、言語知識集合知サーバの導入による用語統一や、産業日本語専用の日英機械翻訳システムの導入など、オーサリング技術以外の基盤技術の開発へ発展させていくことも可能と思われます。また、Japioは米国和文抄録の作成事業も受託しており、こちらは英語から日本語への翻訳ですが、この抄録を産業日本語化することにより、その明晰性や検索精度を向上させていくことも検討すべきでしょう。

また、南特許技監からお話のありました、AIPNにおける審査関連書類の機械翻訳精度向上という課題につきましても、Japioとして産業日本語技術を利用したソリューションを提案させていただければと考えております。こちらにも、強いニーズが存在し、かつ作成者が特許庁審査官に限定されているという点で、産業日本語の導入対象として最適なテーマではないかと考えます。審査関連書類は特許明細書と性質を異にする部分も多く、Japioが自主的に進めております特許明細書を対象とした現在の研究開発ではカバーされない部分もあるかと思いますが、例えば拒絶理由通知書には特許明細書からの引用があり、この引用部分が特に機械翻訳精度向上を必要としている、とも伺っております。少なくともこの引用部分の明晰化にはJapioの研究成果が大いに役立つはずです。また、審査関連書類特有の部分につきましても、過去のノウハウを応用して追加仕様としてまとめることが可能であると考えます。

木下室長からもご指摘がありましたように、不特定多数の特許明細書作成者に産業日本語を普及させるためには、ある程度の強制力をもって使用を推進する活動が必要であると、Japioとしても考えております。具体的には、特許庁にご協力いただき、何らかの指針を出し、それに沿った活動の推進をしていただくことが不可欠と考えます。そのためには、先に挙げた**PAJ作成事業の効率化や審査関連書類の翻訳精度向上等において産業日本語の効果を実証していくことが非常に重要**であると認識しております。

こうして具体的なニーズに即した実用的な開発を積み重ねて実績を挙げ、諸方面からのご協力を賜りつつ、最終的には特許明細書への産業日本語の導入までつなげていきたいと考えます。そして、その後の運用フェーズにてJapioが担うべき重要な任務は、産業日本語の利用基盤を維持するための種々のサービス、例えば特許文書用にカスタマイズした産業日本語仕様（「特許版・産業日本語詳細仕様」）の改訂や、特許業界で共用すべきデータの整備、具体的には用語統一のための共用辞書の作

成・維持管理や、過去に作成された膨大な特許文献を産業日本語化し、特許文書検索用データベースの構築を支援することなど、中立的な立場で実施されるべきメタサービスを事業化していくことでありましょう。

Japioとしましては、このようにして産業日本語プラットフォーム開発に取り組み、産業日本語の導入による特許文書の利用性の向上、ひいては我が国の知的生産性の向上に寄与できればと考えております。確かに、産業日本語という新たなスタンダードを特許業界、さらには産業界全体に普及させていくことは、Japioが独力で達成できるほど容易なものではございません。本座談会にご参加いただいた皆様をはじめ、経済産業省、特許庁及びその関連団体、ベンダー企業や大学・研究機関、そして産業文書の作成者である企業の皆様のご理解とご協力のもと、一致団結して進めることで、大きな潮流にしていく必要がある、と認識しております。本座談会を閉会するにあたり、ご参加いただいた皆様に厚く御礼申し上げますとともに、『産業日本語コンソーシアム』及び『産業日本語フォーラム』等へのご参加を含め、引き続きのご支援とご協力を賜りますよう、お願い申し上げます。

本ネット座談会に参加していただいた研究者のご寄稿も寄稿集にごございますので、ご参照いただければ幸いです。

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)JKAの機械工業振興事業補助金の交付を受けて(財)機械システム振興協会の委託により実施しています。