

H21.8 / 11 ~ H21.10 / 9

ネット座談会

特許分野における機械翻訳の活用と、 特許版・産業日本語への期待

ネット座談会とは、特許情報に関係する各界からの有識者の方々にご参加いただき、メールベースで意見交換をしていただくバーチャルな座談会です。今回は、機械翻訳の活用の実態と将来像、特許版産業日本語への期待をテーマに、約2ヶ月間にも及ぶ時間をかけ、活発な意見交換がなされました。

座談会では、Japio 特許情報研究所長の守屋専務理事による、Japio の取り組みの紹介に引き続き、ご参加いただいた皆様から、本テーマに関する様々なご意見やご指摘等が投稿されました。

参加者の皆様からいただいたご意見等のエッセンスを少しだけ紹介してみたいと思います。

特許庁 南特許技監



★特許審査では、機械翻訳が日本のみならず各国で利用されている。

★特に、「特許審査ハイウェイ」では、機械翻訳が出願人の翻訳コストの負担軽減に寄与している。

日本弁理士会 筒井会長



★特許明細書は技術文書であると共に、法律文書（権利書）でもあり、精密な翻訳が要求される。

★特許明細書やクレームの作成者と、機械翻訳ソフトの開発者とは互いに切磋琢磨しながら、今後のあるべき姿を考え、特許明細書やクレームの作成・研究や機械翻訳の開発・研究を進めていくべき。

日本知的財産協会 中山専務理事



★機械翻訳で完成形をと望むことは、“夢のまた夢”。

★クレーム部分の翻訳問題は、“80点主義”で考えてもいいのではないかと？

キヤノン株式会社 大野副本部長



★特許明細書の特性を認識した上で、機械翻訳の在り方とその利用のあり方を検討すべき。

★現翻訳ソフトがどのような日本語に対して精度の良い翻訳ができるかの評価も不可欠。

富士通株式会社 岩田本部長



★外国特許明細書の質の向上とコスト削減は、待ったなしの状況。

★一発回答方式の頑張りには必要ない。機械翻訳しにくいならば、ユーザーに対してその点を指摘し、日本語の修正を要求してあげればいい。

東京大学 辻井教授



★背景知識を持たない人でも理解できる文にすること、また、解釈が人によって変わらない文にすること、ここに産業日本語の重要な使命がある。

★機械翻訳の基本的な技術が成熟してきたこの時期に、特許を書

く人まで含めたトータルな系で、コストを最小にするシステム構成、明晰な日本語を定義することが重要。

株式会社富士通研究所 潮田主管研究員



★機械翻訳はどのような日本語が苦手か？（訳が文脈に依存する文、長い文、並列表現が多い文）

★権利解釈等で不利益にならない翻訳のためには？—機械は、人間が見落としがちな様々な曖昧性を自ら解消するのではなく

ユーザーに注意を喚起して、可能なオプションをユーザーに提示するという役に徹するべき。

このような“熱いメッセージ”が沢山飛び交う、大変中身の濃いネット座談会です。

さあ、ご興味を持たれた方は、以降に掲載されております、座談会記事を是非ご一読してください。



はじめに Japio 特許情報研究所の取り組み

進行 守屋：8月11日(火) 投稿スタート

皆様お忙しい中、ネット座談会にご参加をいただきまして誠にありがとうございます。私は、進行を務めさせていただきます、Japio 特許情報研究所所長の守屋でございます。

このネット座談会も今年で3回目になります。一昨年の第1回目は、「特許情報研究所設立記念ネット座談会」ということで、ご参加をいただきました皆様から、特許を取り巻く各界の動きについてご意見を伺うとともに、設立間もない Japio 特許情報研究所に期待することについてご議論をしていただきました。その座談会では、「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をいただきました。

昨年第2回目は、「産業日本語特集 ネット座談会」ということで、専門家の皆様のご協力を得て特許情報研究所が進めてきた「産業日本語プラットフォーム開発計画の策定」がメインテーマでした。行政、学界、産業界それぞれのお立場でのご意見や、産業日本語に対する期待についてご議論をしていただき、「産業日本語プラットフォーム開発計画」を具体化していくための考え方を整理させていただきました。

今回は、**特許情報に関する重要な研究テーマの一つである機械翻訳の活用や、機械翻訳の有効活用を進める上**

プロフィール

守屋 敏道

一般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



でのソリューションの一つとして期待される分かりやすい日本語文（特許版産業日本語）などについて、ご議論をしていただきたいと思います。

議論を進めるに当たり、私より、機械翻訳や特許版産業日本語に関する Japio 特許情報研究所の取り組みについて簡単に紹介させていただきます。

～機械翻訳の翻訳精度向上に関する Japio の取り組み～

機械翻訳に関する研究開発は、現在の Japio の事業と大いに関係しております。

Japio が特許庁から請け負っている主要な事業の一つに、特許翻訳に関する事業があります。具体的には、PAJ (Patent Abstracts of Japan : 公開特許公報英文抄録) の作成事業と、米国公開特許明細書、米国特許明細書、欧州公開特許明細書の和文翻訳抄録の作成事業の2つの事業です。

PAJ や和文翻訳抄録の作成事業において、その翻訳品質の維持・向上や作成効率の向上は、これらの事業を遂行する上で常に意識すべき重要な課題と認識しております。

Japio は、これらの課題を解決するために、専門用語辞書の構築に努めてきました。最初は、翻訳の過程で見出した難しい訳や特殊な訳の事例を集めて、それを翻訳者の皆さんで内部共有する仕組みを構築しました。その後、特許関係の対訳事例を本格的に収集するべく、より効率的に対訳事例を収集するための研究開発を進めるとともに、収集した専門用語辞書から機械翻訳用の辞書を作成する研究開発をしてきました。こうして構築された専門用語辞書を翻訳者の皆さんで内部共有するとともに、この専門用語辞書から機械翻訳用の辞書を作り、それを用いて機械翻訳を行った結果を翻訳支援情報として翻訳者の皆さんに提供しています。このような仕組みにより、翻訳品質の維持・向上を図っています。しかしながら、機械翻訳に関する外部サービスを実施する段階には至っておりません。現在は、機械翻訳精度の向上のための研究をさらに深化させ、対訳事例の収集源を拡大す

るとともに、中国語等の他の言語への対応についても研究を始めています。

一方、このような**研究開発を適切かつ効率的に行うためには、機械翻訳に関する研究者や専門家の皆様からアドバイスやご意見をいただきつつ研究を進めることが必要です**。こういった観点から、Japio は、アジア太平洋機械翻訳協会（AAMT）内に、AAMT/Japio 特許翻訳研究会を設置させていただきました。研究会の会長は、この座談会にもご参加いただいている東京大学の辻井教授です。Japio は、この研究会に参加することで、機械翻訳の研究に関する最新情報の収集に努めるとともに、機械翻訳に関する研究者や専門家の人的ネットワークを拡げてきました。

また、Japio は、国立情報学研究所（NII）が主催する NTCIR（NII Test Collection for IR Systems）という、情報検索システム評価用テストコレクション構築プロジェクトに対して、特許検索・特許翻訳タスクのためのテストコレクションとして、日本及び米国の特許文献データの提供を行ってきました。NTCIR の目的は、情報検索やテキスト要約・情報抽出、機械翻訳等のテキスト処理技術に関する研究の効率的な発展を促すことにあります。NTCIR では、テストコレクションの提供や各種コンテスト等を通じ、多くの研究者や技術者が集い、共通のデータやテーマに基づき多様な研究・実験が行われます。特許情報検索や特許情報翻訳は、重要なタスク（研究テーマ）のひとつになっております。

Japio 特許情報研究所では、これらの取り組みにより拡がりつつある人的ネットワークを大事にし、この人的ネットワークを通して得られた知見や研究ニーズを踏まえながら、研究開発を進めております。

～特許版・産業日本語の策定と普及に関する Japio の取り組み～

しかしながら、特許分野における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、特許関連文書の書き方において、何らかの取り組みが必要であることがわかってきました。現在の特許関連文書の問題点として、一文が長く、200 文字を越えるものがざらにあること、係り受けが複雑で読み手にとって分かりづらいこと、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で表現されていること等が挙げられます。

これらの問題点は、機械翻訳の利用性を低下させていることはもちろん、読み手に対して正確にその内容が伝わらないという根本的な問題を生じさせています。

こうした問題に対処するために、Japio は、財団法人機械システム振興協会からの受託事業として、平成 19 年度より産業日本語プラットフォームの開発と普及に関する調査研究に取り組んできました。**特許文書等の産業技術文書の記述に適した「人とコンピュータの双方にとって理解しやすい標準的な日本語（産業日本語）」の策定とその普及システムに関する研究です**。この研究では、**特許明細書を「産業日本語」で記述することにより、以下のような効果が得られることを期待しております**。

- 1) 技術内容等を翻訳者が正確に把握し得ようになるため、誤訳等が少なくなる他、機械翻訳を利用した場合でもその翻訳品質が向上され、外国出願時の翻訳コストを大幅に削減することが可能となる。
- 2) 出願公開される特許情報を利用者が正確に理解でき、また、テキスト検索の効率も向上するため、特許情報共有の促進、検索の効率化が図られる。

平成 20 年度にまとめました「産業日本語プラットフォームの開発計画」について簡単にご説明いたしますと、計画の要となる開発課題は、以下の 4 つの要素から成り立っています。

① 産業日本語共通基盤仕様

産業日本語共通基盤仕様とは、様々な態様の産業日本語に対して、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、その仕様を記述するための共通形式とで構成される産業日本語の基本的な日本語共通仕様です。例えば、特許関連文書の記載に適した日本語の仕様を検討する際には、特許法に規定される記載要件等の特許関連文書の特性を考慮する必要があります。したがって、特許版産業日本語は、このような特性を踏まえ、産業日本語共通基盤仕様を特許関連文書の記述に適するようにカスタマイズしたものに相当します。

② プラットフォームシステム

産業日本語を我が国産業界に広範に普及させるためには、産業技術文書を作成する現場の作成者の負担を軽減する環境を整備する必要があります。「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システムです。「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ます。このオーサリング・システムは、作成対象の文書に適合した産業日本語の仕様に基づく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できます。他方、言語知識集合知サーバは、規則ベースや辞書ベースの内容を集合知によって漸進的に整備していく共通環境です。

③ アプリケーションシステム

コンピュータ処理に対する高い適合性も実現する産業日本語で書かれた文書が利用できることを前提として、各種アプリケーションシステムが開発・提供されることにより、産業技術文書の利用が飛躍的に高まり、我が国の知的生産性の向上につながることを期待されます。開発計画では、文書処理の代表として翻訳と検索を取り上げています。

④ モデル運用サービス

産業日本語共通基盤仕様、プラットフォームシステム、アプリケーションシステムが開発されることにより、産業日本語による文書の作成、その活用といった、産業日本語を円滑に普及させるための基本的な仕組みができあがります。開発計画では、この基本的な仕組みの上に成り立つ二つの「モデル運用サービス」を提案しています。一つは、特許関連文書を対象にした文書の作成、蓄積・検索、変換（翻訳や要約）を高度に支援する知財ワンストップサービスであり、もう一つは、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です。

～特許明細書ライティングマニュアルの作成～

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ります。Japio は、知財分野の専門家の方々や、言語分野の専門家の方々にご協力を得て特許版産業日本語委員会を設置し、産業日本語共通基盤仕様の策定作業を行い、平成 20 年度には、その第 1 版を策定しました。この第 1 版を実用化に向けて更に改良・改訂を重ねていく必要があります。

平成 21 年度も、このネット座談会にご参加いただいている団体や企業の皆様方に協力していただきながら、まず、**特許分野において産業日本語を普及することを当面の目標とし、委員会活動（特許版産業日本語委員会）を継続しております。**実際の特許関連文書の作成現場に産業日本語を導入するための方策について調査研究を行い、その効果を実証しつつ、その適用範囲を広げていくことを考えております。

具体的には、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の仕様のブラッシュアップを行うことを考えております。**特許**

明細書ライティングマニュアルは、特許明細書の内容が理解しやすく、かつ、機械翻訳を有効に活用し得るようになる文章の書き方を、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ます。マニュアル作りを通して得られた情報や知見を活用し、産業日本語共通基盤仕様、特許版産業日本語仕様の改良・改訂を行っていきつ

もりです。
もちろん、このような作業は、Japio 単独ではなかなかし得るものではありません。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し協力しながら作業を進めることにより、はじめてなし得るものです。Japio といたしましては、特許版産業日本語委員会にご参加いただいている皆様のご協力をいただきながら、特許明細書の作成における産業日本語の有効性を実証できればと考えております

以上、機械翻訳や特許版産業日本語に関する Japio 特許情報研究所の取り組みについて簡単に紹介させていただきました。

10 年前に特許庁の特許電子図書館（IPDL）で日本特許公開公報の英語への自動機械翻訳サービスが開始されて以来、特許分野での機械翻訳の利用は国際的にも拡大してきております。しかしながら、機械翻訳は特許分野等での翻訳コスト削減に大きな期待が寄せられているものの、機械翻訳精度の面でまだまだ多くの問題点があり、更に一層の改良・改善が必要であります。

機械翻訳というチャレンジングなテーマを巡って、座談会参加の皆様方から、特許分野における機械翻訳の活用現状やその問題点、機械翻訳活用に向けた解決策等につきまして、ご紹介やご意見をいただきたいと思

います。
特許庁の南特許技監からは、特許審査や特許情報普及における機械翻訳の活用現状や今後の計画等についてご紹介していただき、機械翻訳の課題や、機械翻訳の研究現場への期待、特許版・産業日本語開発の方向性や期待についてお話していただきたいと思

います。
日本知的財産協会の中山専務理事、日本弁理士会の筒井会長、キヤノン株式会社知的財産法務本部の大野副本部長、富士通株式会社知的財産権本部の岩田本部長には、特許制度や特許情報のユーザーとしてお立場から、特許明細書作成時に留意している点、明細書の翻訳問題、外国出願時や外国先行技術調査時等における機械翻訳の活用、明細書ライティングに関する研究等の活動についてご紹介していただき、機械翻訳の課題や、機械翻訳の研究現場への期待、特許版・産業日本語開発の方向性や期待についてお話していただきたいと思

います。
東京大学の辻井教授には、自然言語処理、機械翻訳の研究者のお立場から、特許文書などの機械翻訳精度向上に関する課題とその解決の方向性、特許版・産業日本語開発の方向性や期待について、教授が会長を務めておられます AAMT/Japio 特許翻訳研究会での研究・活動

のご紹介等も交えながらお話していただきたいと思ひます。

株式会社富士通研究所の潮田主管研究員には、機械翻訳システムの開発メーカーのお立場から、機械翻訳の研究・開発現場における機械翻訳精度向上に関する最近の取り組みや、特許版・産業日本語開発の方向性や期待についてお話していただきたいと思ひます。

それでは、皆様よろしくお願ひいたします。

特許審査と機械翻訳の観点から

南 : 8月24日(月) 投稿

それでは、私からは、特許審査と機械翻訳という観点でお話しさせて頂きたいと思ひます。特許審査において機械翻訳が利用される場面として大きく分けて二つあるかと思ひます。一つは審査期間の短縮のためにすでに海外特許庁が行った特許出願の審査関連情報を参照する時であり、二つ目は先行技術文献の調査をする時です。

前者の審査関連情報の参照については、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35以上の国・機関で利用可能となっています。そして、AIPNを利用する国・機関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。なお、本年6月にとりまとめられた知的財産推進計画2009にも掲げられていますが、今後は欧米と比してまだ利用が十分されているとは言えない東南アジア地域等の特許庁を対象にセミナーや研修を行いその利用拡大を図る予定です。

さらに、出願人にとって、高品質の特許の取得、迅速な審査、特許取得コストの削減という三つの大きなメリットを有する施策「特許審査ハイウェイ」においても、AIPNは活用されています。すなわち、日本国特許庁で特許となった出願につき、外国特許庁に対して特許審査ハイウェイを申請する際、日本国特許出願の包袋情報がAIPNによって提供されている場合は、AIPNが提供する機械翻訳文をもって、本来ならば出願人自らが提出するところの包袋情報の翻訳文に替えることができます。唯一、日米間の特許審査ハイウェイは例外であり、日本国特許庁のオフィスアクションとクレームの翻訳文として人手による翻訳文を提出することが現状では求められています。しかしながら、翻訳のコスト負担軽減について多くの出願人から要望がなされていたところ、本



年5月に東京で開催された第2回多国間特許審査ハイウェイ実務者会合において、各国とも、翻訳文のうち、第一庁のオフィスアクションの翻訳文については、原則機械翻訳を受け入れることが合意されました。したがって、今後、米国への特許審査ハイウェイ申請のための書類としてAIPNが提供するオフィスアクションの機械翻訳文を受け入れるよう、日米間のガイドラインの調整がなされることが期待されています。なお、クレームの翻訳文については、クレームの一致性の判断の基礎となることから、機械翻訳の容認には慎重な意見が多かったため、機械翻訳技術の進展を見つつ引き続き検討していくこととなっています。

したがって、さらに出願人の負担軽減を実現するためには、日英機械翻訳の翻訳精度向上が重要になります。現在、AIPNでは精度向上のために翻訳できなかった語(未知語)や欧米特許庁からの誤訳フィードバックをシステムに反映しているところであり、今後も特許庁ができる範囲で機械翻訳の翻訳精度向上に取り組んでいきたいと考えておりますが、機械翻訳技術の研究・開発は主に民間企業や大学等でなされることから、民間企業等の継続的な取り組みに期待をしております。また、仮に将来、産業日本語を適用することにより機械翻訳技術で解決できない点を解決できることになれば翻訳精度の更なる向上に結びつく可能性があるのではないかと期待しています。

一方、機械翻訳が審査で利用される場合の後者として挙げた先行技術文献調査につきましては、これまで特許庁では欧米の特許文献について人手の翻訳により和文

抄録を作成して審査に活用してきましたが、現在、特許庁が進めている業務・システム最適化計画の新検索システムの開発において多言語横断検索システムを導入すべく検討を進めているところです。これは近年、非英語圏の文献として非常に重要となってきた中国、韓国の文献について、機械翻訳を利用して、入力された日本語テキストでこれらの文献を同時に検索して、日本語で表示するものです。この検索システム導入に向けて**昨年度は基礎的な調査を行いました**が、その結果、特に**中国語については韓国語と比較すると現状の翻訳精度では審査に使えるレベルにはまだ達していないことがわかりました**。中国語の機械翻訳を利用した文献検索の実現についても、英語と同様に民間企業等の取り組みに負うところが大きいと考えています。

また、これまでは日本の状況について話して参りましたが、機械翻訳に対する取り組みは日本だけではなく海外特許庁でも整備すべき情報技術関連システムの一つとして注力しているところです。例えば、欧州、韓国、中国の各特許庁でも機械翻訳を利用したサービスを海外特許庁の審査官、あるいは一般ユーザー向けに提供しています。具体的に申し上げますと、欧州特許庁が提供する esp@cenet という日本の特許電子図書館（IPDL）と同様なインターネット・サービスではドイツ語の文献を英語に機械翻訳するサービス等が利用可能です。また、韓国特許庁では日本の AIPN と同様に海外特許庁向けに審査関連情報を英語に機械翻訳して提供しています。

更に、2007年より始まった**日米欧韓中の五大特許庁の協力においては、ワークシェアリングにより審査期間の短縮を図ることを目的としていくつかの情報技術関連プロジェクトがあります**。この中で**機械翻訳もプロジェクトの一つとして取り上げられており**、現在、韓国がリード庁となってその翻訳精度の向上について検討を進めており、我が国も積極的に議論に参加しているところです。

進行 守屋：8月25日（火）投稿

南特許技監、ご投稿をありがとうございました。特許行政において、機械翻訳システムが実際に運用され、日本のみならず各国の特許庁における特許審査を支えていることを紹介していただきました。また、高品質の特許の取得、迅速な審査、特許取得コストの削減という三つの大きなメリットを有する施策である「特許審査ハイウェイ」においても、機械翻訳が活用され、出願人の翻訳コストの負担軽減に寄与しているとお話は、機械検索の関係者にとって勇気づけられるものです。その一方で、機械翻訳精度の更なる向上や多言語対応、特に中国語への対応について、民間企業や大学等での継続的な取り組みへの期待や、機械翻訳技術で解決できない点への解決策としての産業日本語への期待をいただきました。

プロフィール

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



特許明細書の特性を認識した上で、機械翻訳の在り方とその利用のあり方を検討すべき

大野：8月27日（木）投稿

翻訳の品質・スピード・コストの点で、翻訳の能率化の要請が高い現状にあります。その中で、パソコンの高性能化、機械翻訳ソフトの翻訳精度の向上により、機械翻訳が注目されています。

現在、弊社では、英日ソフトを外国特許の翻訳に組織的に活用しています。しかし、あくまでも特許公報記載の概要把握のための機械翻訳であり、英日/日英両ソフトにおいて、正しく訳出するために機械翻訳の精度向上を目指してはいますが、まだまだ未整備な部分や不十分な部分が多いのが現状です。このことを踏まえた上で、日本特許の明細書・クレーム（以後、特別に断らない限り「特許明細書」若しくは「明細書」と記す）の翻訳における機械翻訳の使用についてユーザーの立場から意見を述べたいと思います。

～特許明細書の役割と機械翻訳の適用～

ご存知の通り、特許明細書には、発明を奨励し、産業の発達に寄与するべく、発明の「利用」及び「保護」を図るという二つの役割があります。前者は、発明の技術内容を開示するための技術説明書としての役割、後者は、発明者・出願人が特許権として権利保護を求める範囲（権利範囲）を告示する権利設定書としての役割、です。技術説明書としての役割からは、第三者（当業者）が、その発明を技術的に理解し利用出来るように技術用語を以って発明内容を正確に、出来れば平易に記載する必要があります。他方、権利設定書としての役割からは、権利を求める者は、自分の欲する権利の範囲を自らの意思で特許性を確保しながら確定する傾向が強いという特性があります。そのために、**技術がベースとはいえ**

表現の自由の許す限り、あらゆる表現を駆使して自らが有利になる権利範囲を設定する文章（クレーム）を作成し、その文章で表現された権利範囲を裏付ける（サポート）明細書の文章を作成します。特に、クレームの文章は、公知技術の範囲と求める権利範囲との間に進歩性の溝があるため、その溝の幅をどう設定すれば、特許性を維持して自分の権利範囲を最大限にすることが出来るか、を前提に作成されたものです。将にクレームをどのように表現すればよいかということであり、自ずと技術用語のみならず一般用語も採用して表現を駆使することになります。しかも、**権利設定時や権利設定後に、予期せぬ公知技術が顕在化する場合が多く、公知技術の範囲は変動的であるという視点から、そのための備えの表現もクレーム予備軍や明細書に盛り込むこと**になります。このような状況における文章は、**文章全体だけでなく、文一つ一つにおいても、必ずしも一義的になるものではなく、寧ろ、多義的な場合が多い**ものです。この多義的と不明瞭・不明確・曖昧の意味合いの違いを明確に認識していないと、作成される文・文章は、往々にして**不明瞭、不明確、曖昧に成りがち**です。もっとも、特許法には、権利取得の発明と明細書は、明確に記載しなければならないと規定してある訳ですが。

更に、翻訳上の問題の本質として、言語間における用語同士の意味範囲の違いがあげられます。技術用語、数学用語であれば、言語間での意味の相違はないのですが、その他の用語間では、意味範囲に相違があるのが一般であります。この点を解決しなければ、正確な機械翻訳は本来成り立たないものと考えています。翻訳の宿命ではないでしょうか。

このような特許明細書の特性を認識した上で、機械翻訳の在り方とその利用のあり方を検討すべきではないかと思う訳です。

私は、**機械翻訳は、適用分野に応じて、主力手段で使うか、補助手段で使うか、分かれるのではないかと**思います。特許等の文章を権利文章として権利解釈する場合や文学文章を扱う場合には、**機械翻訳は、将来に渡って補助手段の域を出ないのではないかと**思います。特許の場合も技術文章としてその文章を扱う場合は、**将来の技術発展に伴って、主力手段として使用される可能性が大きくなる**と思います。

企業が特許明細書を利用する目的は、その役割通り技術文献として技術を理解するためであり、権利文献として自他の特許権の権利範囲を確認するためです。

機械翻訳に期待することは、補助手段としての場合と主力手段としての場合とでは、異なるのではないかと、補助手段として求められる機械翻訳と主力手段として求められる機械翻訳に要求されることは、共通なところはあるとしても、相違することも結構あるのではないかと、思います。或いは、特許明細書の分野においては、主力手段として使用されるように構成された機械翻訳システムであれば、補助手段としても十分その役目を果たすシ

ステムになり得るのかもしれませんが。

何れにしても、**より翻訳精度の高い機械翻訳システムの実現には、ルール化やライティングマニュアルの作成が有効に成るのではないかと**思います。その意味で、Japioが進めようとしているプロジェクトは、ある成果を齎し高精度な機械翻訳システム実現の方向性が見えてくるのではないかと期待している訳です。

弊社の場合、現状では、上でも述べたように、権利範囲（クレーム）の部分と明細書中のそれ相当分の記載部分を除いた技術説明の部分、例えば、実施例記載部分や図面の簡単な説明の部分に機械翻訳を適用し、クレームおよびそれ相当の部分に関しては、人間が当初から直接翻訳する、という考えの下で対応しています。

～現状の機械翻訳に求める課題～

人が翻訳する場合の特徴はなんでしょうか。例えば、日本語に頻繁に見られる主語や目的語の欠落を前後の文脈を組みとって適切に補う、等の能力があります。又、長文で分かり難い場合には、意味を汲み取りながら適切な短文に区切って翻訳しようとする。或いは、言語間における用語同士の意味合いの相違を考慮しながらよりその場における適切な翻訳用語を選択しようとする。このような作業行為の結果に拠る**翻訳の精度は、翻訳対象の属する分野に如何に精通し、しかもその精通さに引き摺られる事なく翻訳対象の文・文章の意味するところを如何に客観的に把握出来るかに**拠ると思います。特に、技術をベースとした文章の場合は、翻訳者がその技



術に如何に親密かどうか、その技術分野に親密過ぎて陥りがちな深読みによる誤訳や不適切な意識を如何にしないか、が翻訳精度を決めていると思います。明晰でない文に当たった場合、作者の意を汲んで訳すか、飽く迄も文章上に表出された意を特定して訳すかは、その翻訳の目的によるものとは思いますが、迷うところでもあります。

このようなことが出来る機械翻訳があればと思う訳ですが、現状は遥かに程遠いところにあります。現状の機械翻訳は、人に譬えれば、母体から生まれ出たばかりの赤ん坊ではないでしょうか。しかし、メモリーの大容量化とコスト低下、ソフトも含めたコンピュータ技術の進展に見られるように、技術の進歩は著しく、何れこの課題は解決されるのではないかと思いますし、解決していただきたい事です。私の言でいえば、柔軟な機械翻訳システムの実現です。

ところで、弊社の現状をもう少しお話しすると、弊社は、特許明細書の翻訳の場合、先にも述べましたように、明細書の内容把握に使用しており、その点では現状の翻訳ソフトでも非常に役に立っていると感じています。しかし、曖昧な文や不明瞭な文、或いは、文法上異のある文に当たった場合、とんでもない翻訳になることがあります。このような場合でも、その後、人が翻訳し直す際により手助けになるように直訳される機械翻訳ソフトがあればと思います。

しかし、あらゆるソフトを検証した訳ではないのですが、我々が検証した範囲では、現状の日英翻訳ソフトは、先の様な意味での直訳すら不十分な状態にあると思います。弊社の翻訳現場からは、ユーザーの立場として求める機械翻訳への課題が、2点程あります。

一つは、機械翻訳ソフト自体の改良です。これに関しては、人間にとっての明晰な和文の直訳機能の不十分さです。例えば、「画像データが保存部 207 にあるかどうかに基づいてこの判断は行われる」という、人間にとって明晰な和文を某翻訳ソフトで機械翻訳した結果、”The image data is in preservation part 207 or this judgment is done based on” と訳されてしまいます。

(希望英文:” The determination is performed based on whether image data is stored in storage 207.)。このように、人間にとって明晰な和文に関しては、希望文の様に訳す翻訳ソフトとして開発していただきたいと思えます。

二つ目は、原文に対しての工夫です。もちろん機械の能力にも限界があり、なおかつ元の和文自体が不明瞭・不明確・曖昧である場合もあるため、源流である和文作成時からの改善が必要であることは、云うまでもありません。

これに関しては、次の特許明細書ライティングマニュアルに求める課題の項で述べたいと思います。

プロフィール

大野 茂

キャノン株式会社
知的財産法務本部 副本部長
キャノン株式会社 顧問
1973年学習院大学大学院自然科学研究科
化学修士課程修了。同年(株)キャノン入
社。1989年特許法務本部長。1993年
知的財産法務本部的財産企画センター所
長。2000年知的財産法務本部副本部長。
2003年理事就任。2004年東北大学大
学院工学研究科技術社会システム博士課程修了
(工学博士)。同年専修大学法科大学院客員
教授。2008年顧問に就任し、現在に至る。



～特許明細書ライティングマニュアルに求める課題～

弊社では現在、機械翻訳を本格的な社内営業ベースに持っていくためには、機械翻訳側からのアプローチのみでは限界があるので、その他のアプローチからいろいろな工夫をしています。

その一つが、**明細書の和文の明晰化への取り組み**です。

タームリスト(各案件の明細書中で使用される重要な単語の技術分野別日英対応表)の採用、1文3行ルール(短文化による簡潔明瞭表現)の実行、要注意語のハイライトツール(先行技術に対する不適切表現、文字化けし得る表現、誤訳につながり得る表現)の採用等です。今は、**和文原稿作成基準書(主語を必ず記載する、主語・述語を明確にする等)作成の準備**に入っているところで、これらの一連の取り組みは、今後の機械翻訳の改善にも役立つと考えています。

権利設定書としての翻訳に関しては、先に述べました特許明細書の特性(明細書を書く人の特性が大いにあるかとも思いますが)から、ライティングをルール化するとしても可也の困難があるのではと思っています。そこで、まずは、**特許版産業日本語の開発/特許明細書ライティングマニュアル作成に当っては、技術説明書としての適切な翻訳が出来るようにしていただきたい**と思えます。そのために、**現翻訳ソフトがどのような日本語に対して精度の良い翻訳ができるかの評価も不可欠ではないか**と思えます。

～今後への期待～

日本人の多くは日本語によるコミュニケーションをベースとしているため、世界に向けて情報を発信する際には、翻訳という負荷を必然的に負わなければならない状況にあります。現在では、パソコンの高性能化、機械翻訳ソフトの翻訳精度の向上により、機械翻訳の使用による翻訳もある程度可能です。機械翻訳の発展は、毎年何十万件もの特許が外国に出願されていることから、

特許の機械翻訳ビジネスの需要は大きなものであり、産業への影響力も高いものと思います。それと、機械翻訳ソフトやその周辺技術の開発、それに先立って標準化・ルール化やマニュアル開発に際しては、特許明細書の本来の役割を充分認識した上で、機械翻訳の開発や利用を考えるべきだと思います。

Japio 主催の「産業日本語委員会」は様々な分野の専門家の方々が各自の意見をもちよることができる貴重な場であると考えています。現在、第1回、そして第2回の委員会を終え、徐々に機械翻訳で正しい訳を出すためには何をすればいいかということに焦点が絞られ、機械翻訳ソフト自体の精度の向上が大きな課題となっていると感じます。この委員会を通じ、今後機械翻訳ソフトそのものの精度の向上、そして機械翻訳で正しい訳を出すための明細書ライティングマニュアルの作成を今年度中に何らかの形で完成させられることを期待しています。

機械翻訳に関連した特許明細書作成時の留意事項・課題、日本弁理士会における様々な研究・研修

筒井：8月27日（木）投稿

守屋所長からの機械翻訳や特許版産業日本語に関する Japio 特許情報研究所の取り組みについてのご説明に対し、特許制度や特許情報のユーザーの立場から、また弁理士としての立場から、特許明細書の作成時に留意している点等について紹介させていただきます。



1. 特許明細書作成時に留意している点

弁理士が明細書を作成する時に留意していることは、いわゆる「良い明細書」を作成することであり、今回のテーマである機械翻訳を主眼として作成している訳ではありませんが、「良い明細書」という意味では機械翻訳のテーマと共通する部分もあると思われます。ただ、何を以て「良い明細書」というかについては、諸々意見のあるところであり、機械翻訳に適した明細書が直ちに「良い明細書」という訳でもなく、多面的な検討・判断が必要と思われますので、紙面の制約等の理由から、ここでは機械翻訳が関係すると思われる面のみについて述べます。いずれにしても、我々弁理士は、**明細書作成に当たり、将来、権利侵害を争うときに裁判所で証拠になることも考慮して、発明の保護・活用を最大限に実現できる、当業者レベルの高度な明細書にすることを心掛けて、日々の作成作業を行っている**のであり、機械翻訳を議論するに際しても、そのことを十分理解した上で議論しなければならないと思います。

その様な前提に立って考えてみますと、「良い明細書」の1つの条件は、日本語として理解し易く、分かり易いことである、と考えます。言い換えれば、分かり易い明細書とは、他の言語への翻訳が容易にできる明細書である、とも言えると思います。逆も真なりとは言えないかも知れませんが、折角一生懸命に作成した特許明細書も、表現が分かり難いが故に、翻訳の段階で誤訳されるようでは、発明の保護・活用にとって不利益となりますので、分かり易さは、特許明細書にとって非常に重要な条件であると考えます。

さて、機械翻訳に関連して特許明細書作成時に留意すべき点としては、次のようなものがあると思います。

- (1) 長文は避け、一文をできるだけ短くする。一文は、例えば3行以内の文とし、一文では一つの情報（配置、大きさ、材質、他の部品との係り方、作用、効果など）のみを表現するようにする。ただし、余り短い文章だけに切っても、部品間の相互関係が分かり難くなることもあるので、留意する必要がある。
- (2) 主語、述語、目的語を明確にする。（例えば、主語の後の助詞の次には、できるだけ、句点を入れ、主語が何であるかを分かり易くする。）
- (3) 修飾語はできるだけ短くし、かつ修飾語の掛かり方を分かり易くする（例えば、句点を入れる）。
- (4) 当該技術分野で一般的に使用されている標準的な用語を使用し、特殊な技術用語は使用しない。
- (5) 部品名として、例えば、2つ以上の漢字を繋げた「造語」を使用しない（例えば、「長手方向最外端相当部」）。
- (6) 用語は統一的に使用する。
- (7) 技術的に分かり易い表現を使用し、用語も時代の変遷に合わせる（例えば、今日では「内燃機関」よりも「エンジン」の方が分かり易い）。
- (8) 代名詞はできるだけ使わない（使う場合は、その代名詞が何を指しているかを明確にする）。
- (9) 文章は論理的に書き、起承転結や「5W1H」にも、留意する。
- (10) 明細書全体が流れを持つよう

な順序で説明する。(11) 外国語に存在しない言葉や記号、外国では使用されない単位等は使用しない(例えば、「コ字状」、「ロ字状」のようなカタカナ表現、坪)。

(12) 参照符号等として小文字アルファベットの使用は避け、アルファベットを使用するのであれば大文字にする。(理由:「a」をそのまま機械翻訳すると、不定冠詞“a”と誤解される恐れがある。)(13) 英訳した際に、不可解な文字の羅列となるような参照符号の使用は避ける。(例えば、部品名を省略した数文字以上のアルファベットよりなる参照符号は使用しない。)(14) 日本語は言語として単数形と複数形との区別が明確でない特徴があるので、複数を念頭においているのか、単数を念頭においているのか、を明確にしつつ表現する。したがって、部品を複数設けることが必須の場合には、複数形として翻訳されるよう、日本語の表現に工夫を要する。(15) カタカナ語に対する英訳語が複数存在する場合には、注意する。(例えば、「パス」は、「pass」と「path」のどちらを意図しているのか明確にする。)

(16) 和製英語や、特定の会社で使用されている用語や、現場用語は、使用しないように注意する。(17) 特許請求の範囲の記載は特に分かり難い記載になり易いので、発明構成要素は、箇条書きにするか、要素ごとに分割可能な形式で記載すること等も考える。(18) 能動態と受動態との表現を明確にして記載する。(受動態を多用すると冗長になりがちで、かつ主語が不明確となり易い。)

2. 明細書の翻訳問題

明細書の翻訳問題については、以下のような点があると思います。特に、**特許明細書は技術文書であると共に、法律文書(権利書)でもあり、精密な翻訳が要求されますので、厳格な対応が必要と考えます。**

(1) 技術分野により、同じ日本語でも、互いに違う外国語に翻訳しなければならない場合があるので、注意する。(例えば、「溝」という日本語の場合、構造物の技術分野では、“groove” 又は “gutter” 等の英語になると思われるが、半導体分野では、“trench” とすべき。)(2) 翻訳の用語によって、権利解釈等で不利益になるような翻訳はしない。(例えば、特許請求の範囲における「よりなる」という日本語を、“consisting of” と翻訳すると、米国において権利解釈する時に限定的に解釈される恐れがあるので、要注意。)(3) 特に特許請求の範囲では、ある用語を初出扱いにする(不定冠詞を付ける)のか、既出扱いにする(定冠詞を付ける)のか、注意する。(4) 日本語は言語として単数形と複数形との区別が明確でない特徴があるので、翻訳を行う場合には、複数か、単数か、を明確に意識しつつ翻訳する。(5) 和英翻訳によくありがちな問題としては、不定冠詞の“a”と、定冠詞の“the”との区別が明確になっていない場合があるので、その点にも注意して翻訳する。

プロフィール

筒井 大和

日本弁理士会会長
昭和50年弁理士登録、昭和63年弁理士会常議員、平成7年弁理士会副会長、平成12年弁理士会司法制度対策委員長、平成18年関東経済産業局広域関東圏知的財産戦略本部員、平成19年日本弁理士会執行理事、常議員、関東支部長を兼任、平成21年4月から現職



3. 外国出願時や外国先行技術調査時等における機械翻訳の活用

現在の機械翻訳の精密度等の実情から考えて、特許事務所における翻訳では、機械翻訳の活用は次のような状況と思われます。(1) 外国出願明細書の翻訳のように、精密な翻訳が要求される場合には、機械翻訳は使用しない。(2) 先行技術調査時のように、精密な翻訳が要求されない場合には、(翻訳ソフトを持っていれば、)機械翻訳の利用も考慮する場合がある。いずれにしても、現状では、特許事務所における翻訳業務は、熟練した翻訳者によるものが主流と思います。

4. 明細書ライティングに関する研究等

日本弁理士会では、明細書の記載は弁理士の本来業務の中核を成すものであるという認識の下に、例えば次のような様々な研究・研修を行っています。(1) 活用に有利な特許権を取得できる明細書の作成の研究や、各種の研修の実行(例えば、日本弁理士会中央知的財産研究所における明細書の記載のあり方に関する研究や、同研修所における明細書作成に関する研修等)、(2) 分かり易い明細書の記載のあり方に関する研究や研修の実行(例えば、日本弁理士会特許委員会では、明細書のあり方について、毎年、研究、提言を行っている。)

5. 機械翻訳の課題

機械翻訳の課題としては、様々なものがあると思いますが、例えば、以下のような課題があると考えます。

- (1) できるだけ精度の良い翻訳が可能であること。
- (2) 日本語の特殊性(漢字を使用すること、修飾語が被修飾語の前に来ること等)を克服して、正確な翻訳が可能であること。
- (3) ある程度長い文でも、正確に翻訳できること。
- (4) 修飾語がある程度長くなっても、正確な翻訳ができること。
- (5) 技術分野ごとに異なる訳語を正確に選んで翻訳できること。
- (6) 時代の変化による技術用語の変遷に臨機応変に対応できること。

6. 機械翻訳の研究現場への期待

機械翻訳の研究現場には、次のようなことを期待します。

(1) 時代の変化による技術用語の変遷に臨機応変に対応するため、技術用語の変遷に対応できるソフトを更に改良、発展させること。(2) 上記(1)に関連して、技術用語の変遷に対応する手段として、ネットを利用した翻訳ソフトが既に提案され、実用されているものもあるが、それを更に進化させ、より利用しやすく、かつ、精度の良いものに改良すること。(3) 技術分野ごとに最適な翻訳ができるような機械翻訳技術を研究、開発すること。(4) 日本語の特殊性や、長所も活かしながら、日本語から外国語、あるいはその逆に外国語から日本語への翻訳ができる機械翻訳技術を研究、開発すること。

(5) 技術分野ごとの専門用語辞書を更に充実させること。(6) 技術用語だけではなく、特許等の知的財産の知識も加味し、両者が融合した機械翻訳システムの更なる開発を行い、日本の知的財産業界の発展に寄与すること。

7. 特許版・産業日本語開発の方向性や期待

特許版・産業日本語に対して、次のような点を更に発展させることを期待します。

(1) 産業技術文書、特に特許文書に共通に利用できる産業日本語共通基盤仕様を更に充実させ、実用価値の高いものとする。 (2) オーサリング・システムと言語知識集合知サーバの更なる充実により、プラットフォームシステムを整備し、我が国の産業界において、産業日本語が更に広範囲に実用されるようにすること。

(3) アプリケーションシステムを更に充実させること。(4) 特に特許分野における情報量は膨大であり、世界共通の情報としての利用価値は大きいので、特許分野における産業日本語の利用の充実、および機械翻訳との連動性の強化等について、更に発展させること。(5) 産業日本語の翻訳、特に特許文書の機械翻訳を精度の良いものとし、実用レベルに進化させ、翻訳コストを大幅に削減すること。(6) 特に特許分野における産業日本語の導入を更に進め、現在進行中の特許明細書ライティングマニュアルの更なる充実を図り、実用可能なレベルに進化させること。

機械翻訳で完成形をと望むことは“夢のまた夢”、機械翻訳の精度をあげるという取り組みは“壮大な試み”

中山：8月27日（木）投稿

中山でございます。機械翻訳に関し、ユーザーの立場でというお話でした。ユーザーの立場から夢を語らせていただきたいと思います。



～過去の出来事からの投影～

私が入社して特許部に配属されたとき、特許明細書の仕上げは、外部の印刷屋さんタイプを依頼して納入してもらう（カーボンコピーも合わせ納入）のが普通のルーチンでした。タイプ専門の人もおりました、普段は納入された書類の誤字脱字チェックをしていましたが、緊急な場合に限り“鉛の活字を一個一個拾って”自らタイプを行っていました。

（そうです特許部に鉛活字の群とタイプのセットがあったのです）

特許部の仕事、お役所へ出す書類とはこういうものなんだなぁと感心したことを覚えています。また、カナ釘流の自分の字で社内書類を作成し提出することの恥ずかしさ、何とかできないものかとため息のつき通しでした。あの頃、今日のこと、ワープロで自分が思った通りの文章を自分がタイプできる日が来るとは思いませんでした。

とすれば、技術（ソフトウェア）は、今の不可能（本問では、立派な機械翻訳）を可能に変えるのではないかな。今の夢は、いつかは実現すると期待してもおかしくないはずですよ。

～夢にもレベルがある～

夢の実現にレベルがあるので、ユーザー（コスト意識）という視点で整理して見ます。

一番の望みは、日本語を入力したら“非の打ち所のない機械翻訳”が仕上がってくることでしょう。詳細な説明だけでなくクレーム（特許請求の範囲）部分も。これ

は、クレームという制度があり、その表現の難解さゆえにクレームなのだと思われている限りは、実現不可能と思います。言語の難解さに起因するので、機械翻訳で完成形をと望むことは、“夢のまた夢” かもしれません。ソフトウェアで解決できる課題ではそもそもないのですから。

この夢の実現には、われわれが、クレームに抱くコンセプト、認識を完全に変えることができるのが問われています。平易な表現でクレームを作成しても、権利幅に何の差もないことを誰かが立証してくれる日が来るまで変わらないのでしょうか。実物を使ってこの実験を行ってくれる人、そういう人を見つけれないものなのでしょうか。

ベストがだめなら、次の夢は、機械が、“詳細な説明” 部分だけでも、ある程度完成した形(翻訳)に仕上げてくれることです。それを以って“良” とする。なぜならば、クレーム部分は、現地(米国)の特許弁護士に詳細な説明を読んで作成するよう依頼すれば良いわけで、料金に見合った立派なクレームを作ってくれるはずで、クレームが悪ければ、特許弁護士が技術を理解できなかったためであり(書き手の日本語能力という問題は残りますが)、大いに高額請求に“クレーム”(文句)をつけることができます。

いくつかの山を越えればゴールが見えるのか

この課題は、技術用語のデータベース作りがうまくできれば解決するものかもしれませんが。が事は、われわれが思うほど、簡単ではないようです。Japio では専門用語辞書作りに取り組み始めて来た。しかし、まだ、社内実験どまりということですから。

かなり昔(15年くらい前)、ある米国の特許弁護士から相談に乗ってくれと言われたことがありました。何事かと構えていたら、「ある知財担当から英文特許明細書のチェックを頼まれている。しかし、書かれている内容がさっぱり分からない。本人には強く言えないので困っている。」という話でした。技術がそんなに難しいのかと聞きなおしたら、「技術の問題ではなく、英語で書かれているが、文章になっていない。想像すらできない。」ということでした。機械翻訳のはしりの時代の話です。最近では、かなり改善されてはいるはずで、このような泣き言はないと思いますが。思いたいですね。次に、これも別の米国特許弁護士の話ですが、4~5年前から「日本特許庁の IPDL で検索した内容の英文変換は使えるね。」といってくれます。そのような見方、感想もあるのです。今度は、日本人ですが、「機械翻訳は、使えないわけではないが、使いこなすにはテクニックが必要だ。」ということを書いていました。

この話は、どのレベルの人がどこまで使いこなせるようにするのか(操作性)、分かるようにするのか(機械翻訳の完成度)の問題を提起しています。

プロフィール

中山 喬志

日本知的財産協会
専務理事
株式会社東芝並びに東芝テクノセンター株式会社にて、知財業務に従事(38年間)2009年4月1日現職就任



ユーザーの欲張り、気まぐれ

ユーザーにとっては、機械翻訳は、人手による翻訳コストを下げるために活用したいと思うところが大き、権利行使の際に支障がないものができればそれでいいのです(実はこの要求がきつく、実証できないので困っているのが実情ではないでしょうか)。機械翻訳の改善ステップには、ネイティブスピーカー(その理解力で完成度を知るため)並び特許弁護士(法的に問題ないことの確認で完成度を知るため)の手助けも、必要かもしれません。

ネイティブスピーカーで思い出しましたが、ひところ英国や米国の翻訳会社から日本語が分かる人がいるので品質のよい翻訳ができますよとの誘いがあり、かなり試したことがあります。どうもうまく行ったという話は聞かない。もうやめていますので。品質がさほど変わらなかったからだと思います。

品質は、日本語から来る問題なので、ネイティブスピーカーに近い人で日本語が分かる人がいても解決しうがなかったのでしょうか。それを機械翻訳が超えられるかということになります。そう考えると、**機械翻訳の精度をあげるという取り組みは、“壮大な試み” なのかもしれません。日本人に、日本語を大事に使うようにという教育の肩代わりをしてもらおう計画になるからです。**いや、日本人に、日本語を大事に使うようにという教育は、それはそれで必要で、それを前提とした試みですという答えが返ってくるように思います。

ここでまた疑問が生じます。誰が正しい日本語を、関係者に教育するのでしょうか。正しい教育を行える教育者はいるのでしょうか。どのくらいの訓練を積めば正しい日本語を書けるようになるのでしょうか。この前提が成り立たないと、先に戻りますが、“壮大な試み” になり、道のりは険しいということになります。

ユーザーは、気楽に理想を語ればいいのですが、機械翻訳の完成度UPに携わる人からは、スペックをきちんとしてくれとお叱りをうけそうです。しかし、そこがユーザーで、「人によって、あるいは分野によって、変

わかりますので一概には言えません。」と肩すかしをしそうです。

外国特許明細書の質の向上とコスト削減は、待ったなし

岩田：8月28日（金）投稿

富士通の岩田と申します。特許制度、特許情報のユーザーとしての立場から、日本企業の知財戦略の中に、特許版産業日本語や機械翻訳をどのように位置づけることができるかという観点でお話をいたします。

企業がグローバルに事業を展開するためには、改めていうまでもありませんが、その創造的活動の成果、すなわち知的財産を、外国においても権利として保持することが必要です。特許権をはじめとする知的財産権は、重要な経営資産であって、これを保有することで、事業のグローバルな競争優位性や自由度を確保して、安定した事業収益につなげることができます。

当社グループを例にとると、年間約2500件の発明につき、外国への特許出願を行っています。これは、日本へ特許出願する発明約5000件のうちの、50%にあたります。

そして、外国へ特許出願する発明は、ほぼ全件、米国へ特許出願しています。米国を重視する理由は、米国をビジネスの拠点と位置づけているからです。さらに、米国の特許制度・訴訟制度が特許を権利として活用しやすい仕組みになっているため、米国特許が、事業を守り、事業収益を確保して、グローバルな事業展開を支える要となるからです。

ところで、当社グループは、研究・開発活動のグローバル化を推進していますが、現在のところその拠点の多くは国内です。したがって、日本での研究・開発から生まれた発明を外国に特許出願することが多く、日本語で書かれた日本の特許出願の明細書を、翻訳およびリバイズして行うものが多数を占めています。

翻訳時の誤りを避け手間を少しでも省くために、日本の特許出願の明細書を、外国語へ翻訳しやすい内容にするようさまざまな工夫をしています。例えば、当社では、日本出願明細書を、つぎのような翻訳を意識した規定を含むガイドラインに沿って、作成しています。

- ・ 文の構造を簡潔にし、修飾関係を明確にする。
- ・ 原則3行以内の短い文にする。
- ・ 主語を明確に記載する。
- ・ 代名詞を不用意に使用しない。
- ・ 能動態と受動態を意識的に使い分ける。

発明の技術内容を詳細に記述する「実施例」について

は、このようなガイドラインにしたがって、比較的翻訳しやすい文章を書くことができます。もっとも、翻訳を意識しすぎると、日本語がぎこちなくなってしまう明細書作成の効率が低下しがちであり、この兼ね合いが難しいところです。

一方、特許出願の明細書はそもそも権利書としての役割を担うものですから、発明そのものを記述する部分では、その権利範囲が狭く解釈されることのないように、発明の技術内容を上位概念化した記述が頻出することになります。このため、主語や目的語が省略されたり、係り受けが難解な文になりがちです。

極めつけは「特許請求の範囲」の記述で、ここでは、権利範囲を規定するための構成要素が続出し、構文や係り受けが難解な長文になることが一般的です。

そこで、日本語の明細書を、翻訳用の日本語原稿に書き換えることが一般に行われています。

米国出願用に翻訳するのならば、先ず、明細書全体の枠組みを、日本語のまま、米国出願の様式に合うように組み替えます。

つぎに、比較的翻訳しやすく書かれている「実施例」についても、再度翻訳の観点から見直し、主語の欠落を補充したり、係り受けの明確化を行います。

問題は、「特許請求の範囲」などの難解な日本語の部分ですが、箇条書きを使うなどして、とにかく翻訳しやすい日本語に書き直します。ここでは、日本語の文字は



使っているのですが、もはや日本語の文章にはならないことが多いわけです。このような翻訳用の日本語原稿があつてはじめて、翻訳者はミスなく翻訳を行うことができます。

このように日本出願の明細書をそのまま外国語に翻訳することは、非常に困難なことで、翻訳を職業とする翻訳者にとってさえ、困難を伴う作業です。人間にとっても理解に苦勞する日本語ですから、現在の機械翻訳に頼ることは、なおさら困難といえます。

また、翻訳用の日本語原稿についても、翻訳者にとってはもとの日本出願明細書よりも翻訳しやすいのですが、現在の機械翻訳システムにとってまだ難があることは否めません。

社内の特許翻訳においても、実用に耐えるシステム作りへの協力も兼ねて機械翻訳システムを使っていますが、現状では、特許翻訳者が翻訳を支援するシステムとして使用するまでの実力だと思っています。

以上のような外国明細書作成実務の状況を踏まえると、守屋専務からご紹介、ご説明のありました「特許版産業日本語」と「特許明細書ライティングマニュアル」については、まずは翻訳用の日本語原稿を作成するための道具として活用することが期待できるものと思われる。ぜひ、提供していただけないものかと考えています。

権利書としての日本語明細書を記述でき、かつそれが機械翻訳に耐え得るといふ完璧な「特許版産業日本語」をいきなり完成させることは、素人ながら、言語処理技術の革命的なブレイクスルーが必要で、ここ数年でできることではないのではないかと気がします。

これに対して、翻訳用の日本語原稿は、特許庁へ提出する権利書としての役割は一旦免除されているのですから、これを記述するにあたり、例えば、節・句の係り受けを指示する記号などを含ませることも可能です。「特許版産業日本語」をそのような翻訳ライクな仕様にすれば、それを使用して、オーサリング・システムを介して機械翻訳に耐えうるような翻訳用の日本語原稿を作成することも、遠い夢ではないような気がいたします。

また、機械翻訳ではなく、翻訳者が翻訳するにしても、そのような「特許版産業日本語」で記述された翻訳用の日本語原稿に基づく翻訳の精度は高く、効率も良いと予想できます。

いずれにしても、**日本企業の外国における知財ポジションを向上させ、そのグローバルな事業展開を支援するため、外国特許明細書の質の向上とコスト削減は、待ったなしの状況にあります。**

「特許版産業日本語」と「特許明細書ライティングマニュアル」の、一日も早いリリースを、強く希望します。

プロフィール

岩田 孝

富士通株式会社
知的財産権本部長
兼 富士通テクノロジー株式会社
社代表取締役社長
1971年7月 富士通株式会社
入社以来、広報、海外法務、多
国間にまたがる知財訴訟、大型
仲裁案件を含む知財渉外等の業
務に従事 現在、(社)電子情報
技術産業協会(JEITA)法務・知
的財産権委員会委員長 2008
年6月より現職



また、当社グループでは、当然のことながら米国を中心に世界各国の特許調査を実施する機会が増大しています。

これに伴い、当社の調査専門子会社では、調査グループの全てのパソコンに当社製の機械翻訳システムを組み込んで検索の結果得られる大量の外国特許明細書の内容把握に利用しています。

しかし**外国特許調査は、もっぱら英語での調査が中心になっており、韓国、中国などの特許明細書は調査対象として触れにくいのが実態です。**非英語圏の特許明細書を英語で検索できるサービスも始まっていますが、収録期間や翻訳範囲の点で十分とはいえません。日本語で調査できるサービスを期待したいところですが、機械翻訳の精度は実用の域に達しているとはいえません。特許調査、特許検索は、特許明細書が権利書として高い翻訳精度をもっている必要はなく、技術書あるいはその抄録として利用できる英訳文、あるいは日本語文であれば足りると考えています。まずは、そのような利用を視野に入れたサービスの早期実現を期待しています。このニーズは、非英語圏の特許調査にあって特に高いのではないかと考えています。

機械翻訳の精度向上に向けた機械翻訳の研究・開発現場の取り組み

潮田：8月30日(日)投稿

富士通研究所の潮田と申します。精度向上に向けて、機械翻訳の研究・開発現場ではどのような取り組みが行われているかについて、過去の経緯も交えながらお話ししたいと思います。

機械翻訳の精度向上にはもちろん特許翻訳向けの対処法のように特定の分野や文種に特化した対策も有効ですが、機械翻訳の精度を特許翻訳の現場で十分使えるレベルまで向上させるためには、本質的なところでの技術革新が必要であり、そのための取り組みは機械翻訳の方式



そのものに関わってきます。

機械翻訳の方式には大きく分けて、ルールベース機械翻訳、用例ベース機械翻訳、統計的機械翻訳の3つの流れがありますが、1970年代末の開発初期から現在に至るまで、日本で製品化されている機械翻訳ソフトの殆どはルールベース機械翻訳と呼ばれる翻訳方式が機軸の方式として使われています。ルールベース機械翻訳では、翻訳のための2ヶ国語辞書（翻訳辞書）と、文を解析したり生成したりするための文法ルールが2つの大きなコンポーネントになっており、これらのコンポーネントの良し悪しでかなりの程度まで翻訳精度が決まってきます。翻訳辞書も文法ルールも基本的には人手で1つずつ積み上げていかなければならず開発には多くの時間と労力が必要となります。富士通の翻訳ソフトATLASも、まだまだ精度的には翻訳の現場で十分に満足いただけるレベルに達してはいませんが、それでも現在の精度まで到達できたのは30年におよぶ研究開発の蓄積があったからだと言えます。

このように多分に労働集約的な開発現場ではありますが、1990年代後半に入りコーパスと呼ばれる言語資源が普及するようになると、研究開発の様相も大きく変わっていきました。コーパスとは、一般にテキストの集積を言いますが、特に機械翻訳の研究開発で多く用いられるのは、2ヶ国語の対訳文を集めたパラレルコーパスと呼ばれるものです。大量に対訳文があると、統計処理によりある程度自動で対訳語句を収集することができます。自動で収集した対訳語句を手で品質チェックすることにより、辞書開発の生産性が飛躍的に向上しました。たとえばATLASを例にとりますと、2000年時点で26万語だった基本辞書の語彙数が、2006年には

約十倍の266万語まで拡張されました。機械翻訳において未知語（辞書登録のない語）は機械で処理のしようのない入力であるため、非常に大きな問題である訳ですが、未知語の問題を解決するもっとも確実な方法は辞書の拡充であり、これが過去10年余りの間の機械翻訳の精度向上に大きく貢献してきています。

またパラレルコーパスを用いることで、未知語処理と同様機械翻訳において大きな課題とされていた訳語選択（訳し分け）の問題についても大きく前進することができました。ある単語、たとえば「調整する」という動詞を英語に翻訳する場合、それがエンジンなどの調整の意味で使われている場合は”tune-up”などの訳語が適切ですが、意見や立場などの不一致・相違を調整する場合には、全く違う訳語、たとえば”accommodate”などを選択する必要があります。従来このような訳語選択は人手でルールを作って対応して来ましたが、パラレルコーパスを用いることでどのような単語が近くにある場合は統計的にどの訳語が適切かという識別基準を学習することが可能になり、訳語選択のためのルールの作成効率を格段に向上させることが可能になりました。

この他にも、パラレルコーパスから頻度の高い文や表現を抽出し、その中の可変部分（文によっていろいろ異なった表現を取りうる箇所）を変数化して対訳表現をパターン化することで定型表現の翻訳の精度を向上させるなどのコーパス活用法も生みだされました。

このように、コーパスの活用により、ルールベース機械翻訳の開発効率全体が格段に向上しましたが、実は、このコーパス、特にパラレルコーパスの利用拡大は、少なくとも研究レベルにおいては、上に述べました辞書拡充等の効果をはるかに凌ぐインパクトを機械翻訳開発にもたらしました。それは上にも挙げました機械翻訳方式の1つであります統計的機械翻訳の飛躍的進歩がパラレルコーパスの拡充によりもたらされたことです。統計的機械翻訳とは、パラレルコーパスの解析をもとに構築された統計モデルを用いて翻訳を行う機械翻訳方式で、1949年にWarren Weaverによって初めて提唱され、その後1991年に米国IBMのトーマス・J・ワトソンリサーチセンターの研究グループにより現在の方式の原型が再構築されました。IBMのグループが発表した当初はまだパラレルコーパスの量も種類もごく限られており、また計算機の処理能力も今とは比較にならないほど非力なものだったため、モデルとしての斬新さは注目されましたが、実用的な方式であるとは、少なくとも日本においては受け止められませんでした。また彼らが用いたコーパスはHansard Corpusと呼ばれるカナダの国会議事録を収集した英仏パラレルコーパスであったため、言語学的に近い言語間の翻訳にはある程度使えても日英翻訳には到底使えないだろうという意見も多く聞かれました。しかしながら2000年代に入りますと、ネット上のテキスト量の爆発的増加も後押ししてパラレルコーパスやその他統計処理に有用なコーパスの量が飛

躍的に増大し、かつ世界中の研究者により統計モデルも年々改良され、そしてまた計算機の処理能力も飛躍的に向上したため、統計的機械翻訳は瞬く間に世界の機械翻訳研究開発コミュニティの間で存在感を高め、2005年ぐらいには、「もうすでに統計的機械翻訳は従来のルールベース機械翻訳を完全に凌駕した」という意見が特に欧米の研究開発者の間で多く聞かれるようになりました。

確かに統計的機械翻訳の潜在能力は非常に高いのですが、世界中の研究者がブレークスルーを目指して統計的機械翻訳の研究を推し進めていくうちに、だんだん従来のルールベース機械翻訳が何十年もかけて解決策を模索してきた様々な障壁に同じように突き当たるようになってきました。現在様々な問題が指摘されていますが、中でも最大の問題は、**文の構造を全く無視して純粋に統計処理のみで、文の構造や語順の全く異なる言語間の翻訳を行うことは現実的でないという問題です。その他、動詞、形容詞などの活用や、語尾変化を多く伴う言語の文を統計的機械翻訳で正しく生成するのは容易ではありません。そこで、現在統計的機械翻訳の研究者達はルールベース機械翻訳がこれまで取り組んで来た様々な言語学的課題に新たに取り組んでいます。**

さて富士通の状況はどうかと申しますと、富士通研究所では、もちろん長年ルールベース機械翻訳の開発に取り組んで来ている訳ですが、統計的機械翻訳の潜在能力には早くから着目し、1990年代の中ごろから、少しずつではありますが、ルールベース機械翻訳の枠組みの中に統計処理の成果を組み込む取り組みを行って来ました。特に統計的機械翻訳がルールベース機械翻訳の従来の課題にいわば「回帰」している今こそ、従来の蓄積をテコにルールベース機械翻訳と統計的機械翻訳を融合させたハイブリッド機械翻訳を一步先んじて開発する、ある意味チャンスであると考えたいところです。しかしながら、従来の蓄積がそのままハイブリッド機械翻訳に適用できるかと言いますと、ルールベース機械翻訳と統計的機械翻訳は全く翻訳の枠組みが異なるため、実はここにもまた大きな課題が立ちはだかっています。すなわち、同じ「古い」問題でも、ルールベース機械翻訳で有効だった対処法がそのまま統計的機械翻訳に有効に適用できるわけではないということです。また、統計的機械翻訳の近年の進歩の速度は非常に速く、旧ルールベース機械翻訳の陣営がそれらをいかに有効に取り込んでいくかは、統計的機械翻訳の中に従来のルールベース機械翻訳の成果をいかに取り込んでいくかと同じぐらい難しい問題だと言えます。しかしながら両者の融合は世界中で確実に進行しており、もはや、ルールベース機械翻訳の陣営とか統計的機械翻訳の陣営などという区別は差ほど意味をなさなくなってきました。

ここまでは機械翻訳全般のお話をして来ましたが、ここから特許翻訳のお話を少ししますと、実は近年の日英間の統計的機械翻訳の進展に特許文献が非常に大きな

プロフィール

潮田 明

株式会社富士通研究所 ソフトウェア&ソリューション研究所
主管研究員

1983年 株式会社富士通研究所入社。表面磁気光学効果、空間光変調器、統計自然言語処理、機械翻訳等の研究に従事。マサチューセッツ工科大学修士、カーネギー・メロン大学博士。2007年度から特許産業日本語委員会委員



役割を果たしています。上にも述べましたように、パラレルコーパスの充実が統計的機械翻訳の進展のためには欠かせないのですが、つい最近まで、日英間のパラレルコーパスは、研究コミュニティが自由に使えるものとしては、非常に限られたものしかありませんでした。しかしながら、守屋所長からご紹介がありましたように、Japio が国立情報学研究所の主催する NTCIR のプロジェクトに対して、特許検索・特許翻訳タスクのためのテストコレクションとして、日本及び米国の特許文献データを提供されるようになってから、大量の特許文献日英パラレルコーパスが統計的機械翻訳用に活用できるようになり、この分野の研究開発が大変活性化されました。一昨年から昨年にかけて国立情報学研究所主催で行われました NTCIR-7 の特許翻訳タスク・ワークショップには国内外から 15 の機械翻訳システムがエントリーされ、そのうち 12 のシステムが統計的機械翻訳のシステム、2 つがルールベース機械翻訳のシステム、1 つが用例ベース機械翻訳システムという風に、統計的機械翻訳の盛況ぶりが印象的でした。先ほども述べましたように統計的機械翻訳は日本語と英語のように文の構造や語順の大きく異なる言語間の翻訳はまだ苦手なため、この特許翻訳タスクでも人間による翻訳品質評価ではルールベース機械翻訳の 2 システムに 1 位 2 位を譲りましたが、今後の進展は大いに期待できると思います。

富士通研究所においても特許文献を対象にした機械翻訳の精度向上に古くから取り組んできています。上述しました NTCIR のテストコレクションで用いられたパラレルコーパスは、同じ発明に対して出願された日本語の特許文書と英語の特許文書をつき合わせて統計的に対訳文を抽出する手法を用いて作成されていますが、富士通研究所でも古くから同様の手法を用いて特許文献日英パラレルコーパスを作成し、特許文機械翻訳の精度向上のためのいろいろな研究に活用して来ました。流れは大きく分けて 2 つありまして、1 つはルールベース機械翻訳の枠組みの中での精度向上、もう 1 つは上述しましたハイブリッド機械翻訳の開発です。ルールベース機械翻訳

の精度向上には辞書の拡充も大変有効ですが、**現在力を入れているのは、特許文書の定型性に着目した、定型利用機械翻訳の開発です。**特許文書は、とくに請求項などは1文が長く文構造の解析に失敗することが多いのですが、同時に特許文書特有の表現が多く含まれているために、その定型性がうまく捕らえられると、より正確な文構造解析ができます。文書の定型性を利用するという意味では上述の「対訳表現のパターン化」によるパターン翻訳も同様ですが、この定型利用機械翻訳システムでは単に対訳表現をパターン化するだけでなく、**原文の構造解析の機構の中に定型性を捕らえるためのルールを組み込むことにより、パターン翻訳の正確性と、ルールベース解析の柔軟性を兼ね備えたシステムの構築が可能になっています。**

2つ目のハイブリッド機械翻訳の研究開発ですが、こちらは多少息の長い研究テーマです。世界中で同様のゴールを目指して研究開発が行われていますが、いまだ「決定打」といえる方式が見出されたとはいえません。富士通研究所ではもちろん過去のルールベース機械翻訳の資産を最大限活かせるような形のハイブリッド化を狙っていますが、その際に大きな障害になるのが、フレーズの単位の不一致です。統計的機械翻訳の現在の主流はフレーズベースの統計的機械翻訳と呼ばれるもので、統計処理を行う際に一単語ごとの統計を考えるのではなく、「フレーズ」と呼ばれる複数単語の繋がりを一塊に見立てて統計処理を行う方式です。ところがこの「フレーズ」は、一般に使われている言語学的な意味でのフレーズとは大きく異なり、統計的機械翻訳の枠組みの中で最も有効に活用できる単語の塊で構成されています。ルールベース機械翻訳では言語学的な意味でのフレーズを用いますから、文を構成する基本的な構成要素が食い違ってきてしまう訳です。現在このフレーズの単位の整合性を取ることが、ハイブリッド化達成への近道だと考えて研究開発を進めています。

皆様の熱いメッセージに感謝！

進行 守屋：8月31日（月）投稿

大野副本部長、筒井会長、中山専務理事、岩田本部長、潮田主管研究員ご投稿ありがとうございました。

大野副本部長には、**特許業務の現場における明細書の書き方の留意点、機械翻訳の活用の現状、機械翻訳のあり方等につきまして、詳細かつ貴重なご意見をいただきました。**特に、①特許明細書の技術説明書と権利設定書という二つの役割と特性に着目し、機械翻訳の開発や利用のあり方を検討すべきであること、②機械翻訳は、適用分野に応じて、主力手段として使うか、補助手段として使うのかを考慮すること、③人にとって明晰な和文の直訳機能の向上が機械翻訳システム側に求められる一

方、原文に対しての工夫として、和文作成時からの改善が必要であること、という大変重要なご指摘がありました。また、タームリストの採用、1文3行ルールの実行など、キヤノンでの明細書の和文の明晰化への取り組みを紹介いただきました。そして、特許明細書ライティングマニュアルの作成にあたり、まずは、技術説明書としての適切な機械翻訳ができるようにすべきであり、そのために現行の翻訳ソフトがどのような日本語に対して精度の良い翻訳ができるかの評価が不可欠であると、今後の取り組みの方向もお示しいただきました。

筒井会長には、**特許制度や特許情報のユーザーの立場から、また弁理士としての立場から、機械翻訳に関連して特許明細書の作成時に留意している点や明細書の翻訳問題や特許事務所における機械翻訳の活用等について、簡潔に整理し纏めて頂きました。**「良い明細書」の一つの条件に、日本語として理解し易く、分かり易いことがあげられるとのご指摘は、日本語原文からの見直しの必要性を提起しておられます。特許明細書作成時の留意点として挙げられた、(1)長文は避け、一文はできるだけ短くする、(2)主語、述語、目的語を明確にする、などの18項目は、長年のご経験に基づく大変有益なご指摘であります。また、「明細書の翻訳問題」として、特許明細書は技術文書であり、法律文書（権利書）でもあり、精密な翻訳が要求されるため、①技術分野による用語の訳詞分けや、②権利解釈等で不利益にならない訳語の選択、③時代の変化による技術用語の変遷に臨機応変に対応できる仕組みの充実といったご指摘は、機械翻訳の課題やその研究現場への期待であります。「特許版産業日本語委員会」での検討すべき課題として取り上げ



ていきたいと思えます。

中山専務理事には、特許制度や特許情報のユーザーの立場から、過去の経験を踏まえながらの大なる夢を語っていただきました。昔の特許部では“鉛の活字を一個一個拾って”自らタイプを行っていたとは、若い人には信じてもらえないかもしれません。隔世の感が有ります。また、「機械翻訳で完成形をと望むことは、“夢のまた夢”、機械翻訳の精度をあげるという取り組みは、“壮大な試み”」との見方は、機械翻訳については、ユーザーの欲張りや気まぐれも頭に入れて、鉛のタイプの場合と同様に時間をかけ、知恵を出して取り組んで行くことが必要であると示唆していただいているのだと思います。「技術は、今の不可能を可能に変えうる」ことは、過去の科学技術の進歩が示すとおりであり、我々のモチベーションの源泉でもあります。そして、夢にもレベルがあるのご指摘は、機械翻訳システムの開発やユーザーによる機械翻訳利用のステップについて重要な視点であると考えます。どのレベルで満足するのか重要なご指摘です。また、クレーム記載に関する意識改革、当面の目標の置き方、コスト意識を持って取り組むべきこと、「正しい日本語」とは何か、などの課題をあらためて提起していただきました。

岩田本部長には、特許制度、特許情報のユーザーとしてのお立場から、企業の知財戦略に特許版産業日本語や機械翻訳がどのように位置づけられるかについてお話をいただきました。国内出願の約半数を外国へも出願されている富士通での取り組みのご紹介は、大変参考になるものです。出願明細書の作成において、(1) 文の構造を簡潔にし、修飾関係を明確にすること、(2) 原則3行以内の短い文にすること、等を規定した、翻訳を意識したガイドラインとともに、特許を出願する国の様式に合わせ、翻訳用の日本語原稿に書き換えていることのご紹介をいただきました。そして、「特許版産業日本語」と「特許明細書ライティングマニュアル」については、まずは翻訳用の日本語原稿作成用に活用することが期待できるとのご意見は、産業日本語の開発ステップを検討する上で、十分留意すべきことだと思います。また、ご指摘のあった非英語圏の外国特許調査における機械翻訳活用の実現は、今後ますます拡大が予想される非英語圏での市場形成や市場参入に際し、重要な視点であり、南特許技監からのご指摘にもありますように、民間企業や大学等での継続的な取り組みが必要であると考えます。

潮田主管研究員には、機械翻訳の研究・開発現場ではどのような取り組みが行われているかについて、基本的な機械翻訳の方式や、その研究開発の経緯を交えながら、解りやすく紹介していただきました。現在のレベルに達するまでに30年という年月がかかっており、機械翻訳の精度を特許翻訳の現場で十分使えるレベルまで向

プロフィール

辻井 潤一

東京大学大学院情報学環教授
英国マンチェスター大学教授
国際機械翻訳協会 (IAMT) およびアジア太平洋機械翻訳協会 (AAMT) 前会長、AAMT/Japio 特許翻訳研究会委員長、国際計算言語学会 (ACL) 元会長



上させるためには、本質的な部分での技術革新が必要であるのご意見に、我々の取り組みがチャレンジングなものであることを改めて認識しました。富士通研究所で取り組んでおられる、特許文書の定型性に着目した定型利用機械翻訳の開発や、複数の機械翻訳方式を融合させたハイブリッド機械翻訳が、ブレークスルーを起こすことを期待します。また、2ヶ国語の対訳文を集めたパラレルコーパスが、統計的機械翻訳の飛躍的進歩を促すとともに、効率的な辞書作成や、訳語選択のためのルール作成、定型表現の翻訳精度の向上といった機械翻訳の進歩に大きく貢献してきたことを紹介していただきました。Japio が、NTCIR に提供してきた日本及び米国の特許文献データから、大量の特許文献日英パラレルコーパスが作成され、この分野の研究開発が大変活性化されたとお話は嬉しい限りです。

きちんとした翻訳をするために！

辻井：8月31日（月）投稿

東京大学の辻井です。計算機を使った言語処理、機械翻訳といった研究に、長いものでもう30年ほど、従事しています。この間、1982年から4年間の国家プロジェクト (Mu プロジェクト) で英語・日本語の機械翻訳、また、1989年に京都大学から英国のマンチェスター大学に移るのですが、そこでEUのヨーロッパ言語を対象とした機械翻訳プロジェクト (Eurotra) に3年間、従事しました。現在は、3年前から始まった科学振興調整費の日本語・中国語の機械翻訳プロジェクトに参画しています。このプロジェクトは、5年計画であと2年足らずで、目的は、日中の科学技術論文の翻訳システムです。

このように、30年近く機械翻訳の研究・開発に従事してきました。ただ、30年間という期間、すべてを機械翻訳の研究に従事していたわけではなくて、計算機による言語処理、知識処理といった基礎研究を並行して行って、そこでの研究成果を時々の機械翻訳プロ

ジェクトで試す、というスタンスで研究してきました。

振り返ってみると、言語処理の技術、機械翻訳ともに長足の進歩を遂げたという印象があります。ただ、同時に、この長足の進歩にもかかわらず、機械翻訳というシステムのむつかしさも実感しています。もうすこし正確には、機械翻訳のむつかしさというよりも、人間・機械にかかわらず、翻訳そのもののむつかしさ、といってもよいかと思います。

今ではそういう傾向は少なくなりつつありますが、機械翻訳という言葉を聞くと、何か魔法への期待のようなものが無意識にあって、人間でも難しい翻訳という作業を機械でできるという印象を持ちがちです。でも、たとえば、特許の翻訳というのは、それを専門にするかなり能力の高い人間の翻訳家で行っても、実は相当に難しい作業なのです。

よく翻訳のむつかしさの例として、戦前の日米交渉で China という単語の定義が満州地域を含むかどうかで米国と日本で違ったために、日本が過剰反応して交渉が決裂したとか、ポツダム宣言受託の際に「日本国家統治が連合軍司令官に” subject to” する」の訳を「制限下におかれる」（外務省）、「隷属する」（軍部）のいずれとするかで大激論になったといった例が挙げられます。これらは、翻訳が国の命運を左右したという意味で非常に印象的ですが、同じような問題は、翻訳で日常的に起こることなのです。

現在、日本語・中国語の機械翻訳プロジェクトに従事していると言いましたが、日本語から中国語への特許翻訳では、日本側に翻訳結果を判断する能力がなくて、非常に悪い翻訳が出回っている、という話もよく聞きます。日本語の能力がかなり高い中国人翻訳家でも、専門分野の知識が十分でないと、まったく理解できない、あるいは、ナンセンスな中国語訳を作ってしまう、逆に、中国語も理解できる日本人の分野専門家があまりいないので、翻訳の質がチェックできない、というわけです。翻訳結果がチェックできないという点で、日本語から英語への特許翻訳よりも、もう一段問題が厄介になっている、ということですね。

こういうことを考えると、**きちんとした翻訳をするためには、**

(1) 2つの言語についての十分な能力

(2) テキストが取り扱う対象分野についての知識、

が必要だということになります。また、China や subject-to の例でもわかりますように、これがたぶん一番難しいのでしょうか

(3) テキストを書いた人の意図についての理解

も、実は非常に重要になります。実際、正しい翻訳をつくるためには、翻訳家は技術の専門家や文書（たとえば、特許）を書いた本人に直接意図に関する質問をするといったこと、これに非常に時間がかかる、と言われて

Japio からの支援を受けている AAMT/Japio 特許翻



訳研究会では、如何にして質の高い専門用語の翻訳用辞書を作るか、また、それを Up-to-Date なものにする技術についての議論をかなり集中的に行っています。これは、用語辞書の質が特許の機械翻訳に決定的に重要だということもあるのですが、同時に上の3の知識・技能をいかに人間と機械が分担し、全体としての翻訳コストを下げるかという点からも、専門用語の辞書が非常に重要だと思っているからです。

つまり、2つの言語に関する一般的な翻訳技能は人間が受け持つとしても、人間の翻訳家は、膨大な専門用語に関する翻訳の知識をすべて保持するという点では機械に劣ります。人間の翻訳家が特許を翻訳すると、専門用語の訳を間違ってしまう、結局、何を言っているのかわからない訳を作り出してしまいます。

また、(3)の問題は、機械翻訳が使えるようになるためには、より一層重要です。特許文の中には、人間が読んでもよく意味がわからない文や、いくつもの解釈が考えられてしまう文が頻出します。もとの文が人間でも理解できない場合には、人間であれ機械であれ、良い翻訳ができないことは自明でしょう。専門家が見れば解釈が一意に決まる文でも、機械や技術の非専門家にはいくつも解釈できる文なども、翻訳には大敵であって、翻訳によるコストを大きく上げてしまうことになります。

言語によるコミュニケーションが円滑に進むためには、背景知識を持たない人でも理解できる文になっていること、また、特許など権利関係が重要な文書では、解釈が人によって変わるような文書は困るわけで、ここに産業日本語の重要な使命があると思っています。

進行 守屋：9月2日（水）投稿

辻井教授、ご投稿ありがとうございました。

国家プロジェクト（Muプロジェクト）での英語・日

本語の機械翻訳の研究をはじめとする、様々な機械翻訳に関するプロジェクトへ参画されたご経験を踏まえ、機械翻訳の研究・開発現場での課題、特に機械翻訳の難しさについてご意見をいただきました。機械翻訳の翻訳精度向上に向けた難しさの本質は、翻訳そのものの難しさにあり、きちんとした翻訳をするためには、(1) 2つの言語についての十分な能力、(2) テキストが取り扱う対象分野についての知識、(3) テキストを書いた人の意図についての理解といった3つの知識・技能が必要であるとのこと指摘がありました。そして、課題解決の方向性として、(1) 専門用語の辞書作りが、3つの知識・技能をいかに人間と機械が分担し、全体としての翻訳コストを下げるかという点からも非常に重要であり、

(2) 言語によるコミュニケーションが円滑に進むためには、背景知識を持たない人でも理解できる文になっていることが必要であり、また、特許など権利関係が重要な文書では、解釈が人によって変わるような文書は困るわけで、ここに産業日本語の重要な使命があるとのこと指摘は、今後の取り組みの具体化にあたり、大変貴重なご指摘であると考えます。

2 巡目のご意見・ご紹介について

進行 守屋：9月15日(火) 投稿

特許分野における機械翻訳の活用と特許版産業日本語への期待について、Japio 特許情報研究所での取り組みの紹介を皮切りに、皆様方の取り組みをご紹介いただくとともに、今後の機械翻訳の研究開発や特許版産業日本語に対するご意見やご期待を多数いただきました。

これまでにいただきましたご指摘等を踏まえまして、機械翻訳をめぐって更にご意見等を伺いたい点について私より提示させていただきたいと思っております。

(1) 特許審査における機械翻訳システムの活用と出願人の翻訳コストの削減について

特許審査において、機械翻訳システムが実際に運用され、日本のみならず各国の特許庁における特許審査を支えており、特に、「特許審査ハイウェイ」においては出願人の翻訳コストの負担軽減に寄与していること、出願人の負担軽減を進めるためには、日英機械翻訳の翻訳精度向上が重要であり、翻訳できなかった語(未知語)や欧米特許庁からの誤訳フィードバックをシステムに反映していること、を南特許技監から紹介していただきました。

「特許審査ハイウェイ」や審査結果の他庁での利用を推進し、翻訳負担を軽減する上で機械翻訳の精度向上は重要なことでもあります。そのために、機械翻訳の元になる審査官の起案文書の工夫やそれに対応した機械翻訳システムの一層の改善も望まれると思っております。この点につ

いて、特許庁の今後の取り組み等について、南特許技監にお考えをお聞かせ願いたいと思っております。

(2) わかりやすい明細書の作成や機械翻訳を活用するための取り組みについて

わかりやすい明細書を作成し、あるいは、的確な翻訳結果を人手あるいは機械により得るために、ユーザーの皆様方は、既に色々な工夫をされ、それらの工夫を基準書やガイドラインの形でとりまとめ、あるいは、研修や研究会を通して、文書作成の現場で共有するといった取り組みを行っておられることを、紹介していただきました。特に、大野副本部長や岩田副本部長には、企業の現場での取り組みについて詳しく紹介していただきました。

そこで、大野副本部長や岩田副本部長には、企業の現場でのこれまでのご経験を踏まえて、さらに何に取組みればよいのか、あるいは、何が足りないのかといった課題とその解決のために今後の取り組むべき方向性について、少し具体的にご意見をいただけますと幸いです。併せて、「特許審査ハイウェイ」等における機械翻訳を活用した翻訳負担の軽減のための特許庁の取り組みに対し、制度ユーザーとして期待することについても、お聞かせ願えますでしょうか？

また、わかりやすい明細書や機械翻訳しやすい明細書を書くにはどうすれば良いか、さらには、不可侵領域とされる(!?) 特許請求の範囲の記載を明晰にするにはどのような工夫が考えられるのか等の問題について、知財業界全体として取り組みをしていくことも、特許の国際戦略が重要となる状況のなかで価値のあることではないかと思っております。簡井会長、中山専務理事には、明細書や特許請求の範囲の明晰化について日本弁理士会や日本知的財産協会での委員会や研修会等で検討してみること等、ある種チャレンジングな取り組みの可能性や問題点について、ご意見をお聞かせ願いたいと思っております。

(3) 特許制度ユーザーのニーズに対応する機械翻訳システムの研究開発動向について

特許分野においては、現に機械翻訳システムが活用されており、機械翻訳システムに対するユーザー側の皆さんからの期待や要望は大きなものがありました。簡単に整理しますと、①現行の機械翻訳システムはどのような日本語に対して精度の良い翻訳ができるのかをしっかりと評価すること ②まずは技術説明部分に対する翻訳精度を上げ、少なくとも人にとって明晰な日本語文の直訳機能を向上させる等、コスト意識を持ちつつ段階を追って進めていくこと ③技術分野による用語の訳し分け、権利解釈等で不利益にならない訳語選択、時代の変化による技術用語の変遷への柔軟な対応、翻訳に必要な知識や技能の人と機械との間での最適分担といった機能を実現するべく専門用語の辞書構築を行うこと ④中国語等の非英語圏の言語に対する機械翻訳の精度を向上させるこ

と等があげられております。

そこで、辻井教授には、こういった機械翻訳システムに対するユーザー側からの期待や要望に関連した、学界での議論や国家プロジェクト等の動きについて、お聞かせ願いたいと思います。

また、潮田主管研究員には、ユーザー側からの期待や要望に関連する機械翻訳の研究・開発現場での現在の取り組みの状況と今後の方向性等について、ご紹介していただきたいと思います。どうぞよろしく願います。

翻訳の難しさについての具体例

大野：9月29日（火）投稿

前回の投稿でも一部述べさせていただきましたが、翻訳の難しさについて具体例を交えて更に説明（強調）させていただきますと思います。

日本においては、特許法施行規則にあるように、「技術用語は、学術用語を使用すること」、特別に定義しない限り、「用語は、その有する普通の意味で用いること」、となっています。しかし、実際の明細書においては、確かに学術用語は使用されていますが、明細書の中は大半が、一般用語で記載されているのが実情です。

学術用語ならば、技術的意味は明白だといえますし、言語間においても技術的に統一されています。しかし、一般用語だと一つの言語でも意味合いに幅があり、それが翻訳の難しさの起因になっています。

次に実例を2つあげてみます。

1つ目は、「振れ形状を有する部品」を表現するのに、「振れた形状の部品」と明細書の中はもとよりクレームにもその表現を用いて記載した出願での翻訳上起きた実例です。この出願での「振れた形状」というのは、振れ餅、餅を振り切る、とか言う場合の餅の振れ形状に似た形状のことです。因みに物理学用語の「振れ」は、一端を固定した柱状の物体に、中心軸を軸とする偶力が加えられたとき、その物体に起こる変形、と定義されています。

該当の出願においては、日英翻訳の場合は、「振れた」⇒「twisted」と翻訳されました。しかし、英韓翻訳では、「twisted」⇒「??(ウエゴク)」と訳されたのです。この韓国語「??(ウエゴク)」は、日本語だと「歪曲した」になるそうです。この違いは、何処から生じたのでしょうか。日本語の「振れた」を和英辞書（新和英大辞典第5版：研究社）で引くと、1.（ねじ曲がった）「twisted」「distorted」「warped」「buckled」；（植）「tortuous」、2.（ひねくれた）「perverse」「cross-grained」「contrary」といった語が記載されています。この「2」では、不適切であることには、異論がないものと思います。他の辞書を見ると「crooked」「gnarled」といった単語も見受けられます。

同じ辞書で「振れ」について調べてみると、「twist」「wrench」「distortion」；「（物）torsion」「（樹木の幹の）intorsion」、といった語が記載されています。

これらの語を今度は、逆に英和辞書（新英和大辞典第6版：研究社）で引いてみると、以下のようになっています。原型が動詞の場合は、動詞で見えます。

「twist」：よる、より合わせる、編む、巻く、振る、ひねる、ゆがめる、ねじる、歪曲する、

「distort」：ゆがめる、ねじる、曲げる、曲解する、歪ます、

「crook」：曲げる、湾曲させる（する、）

「tortuous」：ねじれた、曲がりくねった、遠まわしの、回りくどい、

「warp」：そらせる（そる）、ひずませる（ひずむ）、曲げる、ゆがめる（ゆがむ）偏る、

「gnarl」：ふしくれたつ、ねじれる、ごつごつする、

「gnarl」は、こぶし、こぶなどに起因する言い回しの際に用いるものですし、「crook」は、腕、足などの場合に用いられ、「tortuous」は、主に言い回しに用いられているようです。「distort」の技術用語は、電子工学分野、又は光学分野で用いられています。

これらからすると、該当の出願の「振れた」の英訳は「twisted」が妥当であることになり、出願当初の日英翻訳に誤りはなかったといえます。しかし、韓国語に翻訳する際に翻訳担当者が「twisted」を受け身態の「be twisted」として解釈し、その原型である他動詞の「twist」を英韓辞書で調べたようです。英韓辞書によると、他動詞の「twist」の対応韓国語は、①「搓る」、②「巻く」、③「顔を歪める」、④「歪曲する」、⑤「ねじる」が記載されています。韓国の翻訳担当者は、辞書において上位の意味になる「歪曲される」に翻訳したということです。「歪曲された」と「振れた」では、意味合いが違います。意味の幅からすれば、「歪曲した」>「振れた」の関係になり、「歪曲した」は「振れた」の上位概念になっていますので、何が問題なのだという人がいるかもしれません。

一般に、特許関係の翻訳者は、狭い意味に翻訳することを嫌い広い意味に翻訳する傾向があります。ここに、3つの問題が生じる可能性があります。問題の一つは、優先権主張して韓国に出願しているとしたら、優先権を消失している可能性があるということです。もう一つは、サポート要件違反に成りかねないということです。3つ目は、先行技術との差（特許性）が、難しくなってくる可能性があることです。この出願の例の場合は、3番目に当たります。

2つ目は、インクジェット記録技術に関する出願の例です。

その出願では、インクジェット記録装置を、インクに限らず色々な液体を噴射して記録することができるという意味で「液体噴射記録装置」と記載してあります。翻

訳において、この「噴射」の英訳が問題になった例です。「噴射する」の英訳には、先ほどの新和英大辞典によれば、「jet」「spray」「inject」「emit」が記載されており、他の辞書には、更に「eject」が記載されています。この例の場合、「噴射する」を「inject」と一貫して翻訳されてしまったのですが、実際に、この翻訳ミスが発見されたのは、特許が発行されいざ提訴せんとした時です。それまで、誰も気付く者がいなかったのは、これまで何十件も翻訳されているし、間違いもなかったという過信があったからかも知れないのですが。ただ、発覚後、原因を探ったところ、特許事務所の翻訳者が、それまでとは別の人になったことだと分かりましたが、この翻訳ミスは、100件程に及びました。不幸なことはその翻訳者が、車のエンジン分野の翻訳が多かったということです。「inject」は閉じられた空間に噴射する意味だし、「eject」は開放空間に噴射する意味であることは、お分かりかと思います。

いずれにしても、特許の翻訳においては、技術の理解、発明の理解があって初めて翻訳であるということでしょうか。しかも、技術・発明の理解が出来ても、言葉の意味に広がりがあり、その広がりが言語間で異なっているばかりか、辞書におけるその言語の意味の記載順序も言語によって異なる現実がある以上、執筆者の意通りに特許明細書の各文を訳すことは、難しいと思います。

しかしながら、各言語の単語の持つ意味の範囲に幅があるのが事実ではあるものの、それとは別にある程度の業界での対訳の統一は必要であり、価値のあることではないでしょうか。弊社でも過去の各案件における単語の蓄積データを整理し、ある程度の対訳の統一を図ることを目的として、ひとつの共通辞書を作成しています。意味に応じて訳語を使い分ける必要がある単語に関して

は、それぞれの訳語をどのような状況で使うべきか、例文を挙げて説明してあります。訳語の統一ができる単語に関してだけでも統一することによって、間違った訳語の使用を極力減らすことができると考えています。これは、弊社での共通辞書ですが、業界全体の底上げのためにも、同様な考えで業界でも統一した辞書を作成することは価値あることではないでしょうか。

このように単語レベルでも難しさがありますが、文単位となると、様々な表現の仕方・書き方が加わることで、より一層の翻訳の難しさがあります。しかし、同じ文意ならよりわかりやすい表現・書き方に統一するなど、文の表現・書き方を統一できるところは統一することで、機械翻訳にも対応しやすくなる文章になるのではないのでしょうか。そのためのルール作りは必要になってくると思います。

しかし、機械翻訳精度を高めようとするればすれ程、対応するためのルールは細かく膨大なものになるでしょう。そして、現実問題として、その膨大なルールを執筆者のすべてに徹底させることには限界があると考えていますが、出来るだけそのルールに従って文章を書いてもらえば、それはそれで、その機械翻訳は自動翻訳により近づく補助ツールとして役に立つものと思います。特に、科学論文などの純粋な技術文章であれば、ルールをそれ程細かく膨大なものにしなくても、自動翻訳が早晚実現するのではないのでしょうか。他方、特許の明細書のように、技術文章と法律文章の性格を合わせもつ文章の場合は、更に工夫する必要があると思います。そのひとつが翻訳パターンの採用ではないかと思います。

現在の翻訳ソフトは、特許専門の翻訳ソフトにおいても、特許特有の翻訳パターンが欠如しているように思えます。特許特有の翻訳パターンを多数登録することにより、日本語の構造を極力損なわないような翻訳に近づけるのではないのでしょうか。つまり、特許明細書での特有のいい回しや表現に関しての翻訳パターンをより多く翻訳ソフトに組み込むのです。類似のいい回しや表現に関して体系化しその体系化に基いた翻訳パターンを用意します。そして、先ず、翻訳すべき和文が、用意されている和文例のどれに構造的に近いかを分析し、一番近い和文例の翻訳パターンに沿って翻訳するシステムにします。そして、その翻訳の妥当性について人がチェックし、若し、修整が必要なら修整してその和文と翻訳文を新しい翻訳パターンとして新たにシステムに登録する、といった学習機能も設けます。更には、執筆者による文章癖も体系化してシステムに覚えさせれば、より精度の高い翻訳システムが出来るのではないかと思います。

更に、日本語の場合、一つの文が複数の意味に解釈できる場合がよくあります。そのため、前後の文脈からその意味が正しいのか判断し適切に翻訳する機能がシステムには必要となってきます。この場合も、それぞれの意味に対応した翻訳パターンをシステムに登録して置き、前後の文脈と整合する翻訳パターンを選択して翻訳する



仕組みにすれば、機械翻訳もより自動化に近づくのではないかと思います。

ここまで述べてきましたように、特許明細書の完全日英自動翻訳システムの完成までの道程は遠いものですが、携帯版自動同時通訳システムにも繋がる開発ですので、ソフト開発者、システム開発者、言語学者、特許専門家等、各分野の専門家が一致協力して一步一步確実に前進させることが大事だと思います。そして、専門同士の境界で不整合が起こらないように、また、現象に囚われて問題解決の本来のところを見失ったりしないように、相手の専門領域もある程度理解する必要はあると思います。そうすれば、開発のスピードも上がるのではないのでしょうか。ちょっと、飛躍しますが、自動同時通訳システムにも繋がる自動翻訳システムの開発は、バベルの塔建設前のように、世界の人類が一つの言語で何不自由なくコミュニケーションを交わしていた時代の到来を夢見ることができることにもなるでしょう。

現在の機械翻訳においては、執筆者の意の通りに特許明細書の各文を正確に訳すのは難しい状況にありますが、先ほども述べましたように、技術説明の部分であれば、日本語を省略することなく構文に従ってきちんと書くという前提で、精度の高い自動化が比較的早めに実現するのではと期待しております。

それまでの間に是非実現して頂くこととして、ユーザーとしての翻訳現場から一つ提案させて頂きたいと思っています。

先ほどもいいましたが、現在では、機械翻訳に100%頼ることができず、補助手段としての活用に留まっています。そのため、機械翻訳された文章を人為的にチェック・修整する作業が必要になりますが、このチェック・修整作業の効率が悪いのが現状です。場合によっては、その技術分野に精通した翻訳者が初めから翻訳した方が遥かに効率がよい、ということがあります。しかし、いつでもそのような翻訳者をあてることが出来るわけではありませんので、結局は、機械翻訳して人為的にチェック・修整するか、精通した翻訳者の手が空くまで仕事を棚上げするか、いずれにしても非効率な仕事になります。

そこで、機械翻訳の効率を上げるために、そのチェック・修正作業の時間を極力少なくするための新たな便利ツールの開発を望んでいます。ひとつ例を挙げさせていただくと、和文の一文単位毎のワード数カウント機能です。現在の日英翻訳ソフトにおいては、和文が長文になればなるほど正しく翻訳されることが少なくなり、訳文を解読し修整する時間が累乗的にかかります。そこで、○○文字以上の和文は訳文を参考にせずに人為的に一から訳し、○○文字未満の和文の訳文だけチェック・修整することができるように、一文単位毎のワード数のカウント機能を設けていただきたいのです。現在は、一文単

位毎のワード数を人為的にカウントしていますが、このやり方でも、経験上翻訳効率を上げているので、一文単位毎のワード数を自動カウントする機能を有する翻訳ソフトの開発は、大いに期待しているところです。そうなれば、機械翻訳を使用する翻訳業務の効率も上がり、機械翻訳の発展にもなるのではないのでしょうか。

日本語の歴史を振り返ってみますと、グローバル化する経済社会は、事実上の世界標準言語といえる英語でのコミュニケーションをベースに動いており、日本語によるコミュニケーションをベースとするわが国産業界にとって、世界への情報発信がビハインドしているのが現状であります。だからといって、英語を日常語化してコミュニケーション言語が英語になれば、日本語の良さや・美しさ（良く英語の文章の中にローマ字で日本語が記載されることがあるが、日本語の良さは、発音のみで、形や構文の良さや美しさは皆無である）を失うことになり、日本文化、畢竟は、日本国を失うことになりかねません。

しかし、わが国においては、言語の相違に基づくコスト負担が大きいことは否めず、また、情報の発信や理解のスピードにおいても遅れをとり、そのことが、産業力、経済力、市場支配力などの国際競争力に影響しているのも現実です。

これらの状況を打破する一つの手段が、高精度自動翻訳システムであると考えています。自動翻訳システム実現の暁には、日本の経済力は、世界のトップレベルになるものと期待します。

高精度自動翻訳システムが実現するまでは、開発途上でそれぞれに開発されたシステムを、大いに使用しその効果をアピールすべきでしょう。その一つの実例が、「特許審査ハイウェイ」等における機械翻訳の活用ではないのでしょうか。

いずれにしても、**高精度自動翻訳システムが実現するまでは、開発途上でそれぞれに開発されたシステムを、用途限定して使用する、活用法を工夫して使用する、或いは一部カスタマイズして使用する、などして、開発者もユーザーも機械翻訳を市場に弛まず浸透させ続けることが、高精度自動翻訳システムの実現を早めるもの**と思います。

今後も色々な角度から検討・議論を！

筒井：9月29日（火）投稿

守屋所長からのご要望によれば、本座談会の二巡目における日本弁理士会への検討依頼テーマとして、「明細書や特許請求の範囲の明晰化について日本弁理士会の委員会や研修会等で検討してみる等、ある種チャレンジな取り組みの可能性や問題点について、ご意見をお聞かせ願いたい」とのことですので、それにつ

いて以下のように意見を述べさせていただきます。

1. 検討の前提

このテーマを検討するに当たり、その前提として、「わかりやすい明細書や機械翻訳しやすい明細書を書くにはどうすれば良いか、さらには、不可侵領域とされる(?!)特許請求の範囲の記載を明晰にするにはどのような工夫が考えられるのか等の問題について、知財業界全体として取り組みをしていくことも、特許の国際戦略が重要となる状況のなかで価値のあることではないかと思えます。」という守屋所長のご提言に対し、色々なスタンスがあり得るのかも知れませんが、ここでは、守屋所長のご提言に沿って本テーマを検討することに致します。

2. 明晰化への取り組みの可能性

まず、今回のテーマのうち、「明細書や特許請求の範囲の明晰化について日本弁理士会の委員会や研修会等で検討してみる等、ある種チャレンジングな取り組みの可能性」については、最終的な結論は現状では見えておりませんが、「取り組みの可能性はある」という回答をさせていただきます。

この点に関しては、一巡目の意見においても紹介させて頂いておりますように、日本弁理士会では、明細書の記載は弁理士の本来業務の中核を成すものであるという認識の下に、これまでも、例えば、活用しに有利な特許権を取得できる明細書の作成の研究や各種の研修の他に、分かり易い明細書の記載のあり方に関する研究や研修を、例えば日本弁理士会中央知財研究所や特許委員会等で実行し、また明細書のあり方について、研究や提言を行っております。

3. 特許請求の範囲(クレーム)の明晰化について

次に、「不可侵領域とされる(?!)特許請求の範囲の記載を明晰にするにはどのような工夫が考えられるのか等の問題」については、更に難しい検討が必要と思えます。なぜならば、**特許請求の範囲(クレーム)の記載のあり方はそれを作成する弁理士個人の研究や実践的経験から得られた「クレーム作成哲学」等も色濃く影響する他、その前提として、依頼者の発明を如何に最良の結果が得られるようにクレームを作成するか、という弁理士にとって最も重要な職業的要請が厳然として存在するから**です。

つまり、後者の条件を最優先すれば、「明細書や特許請求の範囲の明晰化そのものには、必ずしも反対というわけではないが、明細書や特許請求の範囲の技術書或いは権利書としてのそれぞれの機能からして、第一義的には、出願人(発明者)の発明がきちんと保護されることが最も重要なことであり、明晰化や翻訳のし易さが最優先課題ではない」ということにならざるを得ないでしょう。



それに対し、「しかし、そうは言っても、明細書や特許請求の範囲の明晰化も重要なことであり、弁理士としても、色々な議論や工夫をして、より分かり易い明細書や特許請求の範囲を作成する努力はしなければならない」というスタンスに立てば、それを前提として、そのためにはどのような明細書や特許請求の範囲の記載のあり方が最適かという方向での検討に入ることができると思います。

そこで、もう一歩進んで本テーマを考えてみますと、「わかりやすい明細書や機械翻訳しやすい明細書を書くにはどうすれば良いかという命題」と、「良い明細書および特許請求の範囲の記載のあり方はどうあるべきかという命題」とは、互いに相反するもので、どこまでも平行線で、決して交わらない乃至収斂しないものなのでしょう(?!)。これらの2つの命題は、必ずしも互いに相反するものではなく、作成者はその調和点を求めて努力をする必要があると改めて思います。

そもそも、弁理士が明細書および特許請求の範囲を作成するに当たっては、「分かり易い明確な明細書および特許請求の範囲」を作成するよう、当然のように配慮しているところです。それと共に、明細書および特許請求の範囲(クレーム)作成に当たって留意すべき重要な事項は、「発明の本質を捉えて、先行技術との差異を明確にしつつ、無用な限定のない明細書および特許請求の範囲(クレーム)が書けているか」ということです。すなわち、「良い明細書およびクレームか否か」ということが極めて重要で、しかも非常に難しく、多大なエネルギーを要するポイントです。そして、その結果が、機械翻訳に適した内容になっていれば、より素晴らしいことです。しかし、それは、必ずしも作成者の最終目的として頭に描かれている最大のポイントではありません。作

成者は、明確で、分かり易く、且つ、良質の明細書およびクレームを作成することに、全神経を集中して努力しています。その努力の結果として明細書およびクレームが出来上がったとき、それが機械翻訳に適していないと判断されるようなものであったならば、誠に不幸なことです。さりとて簡単に書き直せば済むというものでもありません。なぜならば、それは、作成者にとって散々検討してやっと出来上がった究極の明細書およびクレームだからです。

このように、特許請求の範囲（クレーム）の明晰化や翻訳のし易さというテーマは、ユーザー、弁理士、言語学者、機械翻訳ソフト開発者のみならず、知財業界全体として今後も色々な角度から検討および議論して行かなければならない課題であると考えます。

4. まとめ

諸々検討していて考えさせられることは、特許明細書、特にクレームは、最も機械翻訳が難しい文書の代表例の1つではないかということです。それは、特許明細書は技術書であると共に権利書であるという側面も併有しており、特にクレームは権利書そのものであるということが原因と思われる。このことは、特許明細書およびクレームの法的性格が変わらない限り、今後も同じと思われる。将来、特許明細書、特にクレームの機械翻訳が簡単に精度良く行えるようになったとき、他の領域に於ける殆ど全ての文章は、機械翻訳で簡単にOKという時代になっているであろうとの意見もある程です。

日本語は「文化」、発明は「文化」を発展させる「技術的思想」であるとするならば、この「技術的思想」を日本語で表現する特許明細書およびクレームの作成も一つの「文化」だという言い方ができます。一方、特許翻訳は、既に日本語として特定された明細書およびクレームを、異なる文化によって生まれ育った別の言語の明細書およびクレームに変えることです。したがって、機械翻訳の開発に当たっては、それぞれの言語の拠って立つ文化が互いに異なるものであることを意識して、より良い機械翻訳のあり方を追究して頂ければ幸甚です。

究極的には、「日本語で表現される特許明細書およびクレームの作成の文化」と、「機械翻訳で目指す翻訳先の他の言語の文化」とのマッチングがとれることが理想ではないでしょうか。そのためには、日本語で特許明細書およびクレームの作成を行う側の人と、機械翻訳ソフトを開発する側の人とが互いに切磋琢磨しながら、今後のあるべき姿を考えて、特許明細書およびクレームの作成・研究並びに機械翻訳の開発・研究を進めて行くことが必要と思います。

言い換えれば、特許の国際戦略の重要性が叫ばれている今日および将来において、知財業界全体として本テーマに取り組んで行くことは、価値があるものと考えます。

特許庁の機械翻訳の精度向上等の取り組み

南 : 9月30日(水) 投稿

それでは、機械翻訳精度向上のための機械翻訳の対象となる審査官の起案文書の工夫や機械翻訳システムの改善についてお話したいと思います。

まず、機械翻訳の対象となる審査官の起案文書の中で、審査官が出願人とのやりとりで作成する主要な書類の一つとして拒絶理由通知書が挙げられますが、先ほど申し上げましたAIPNではこの拒絶理由通知書を日英機械翻訳システムを介して海外特許庁に提供しております。

当然のことではありますが、拒絶理由通知書は出願人が理解しやすい文章で記載することが一番重要であると考えています。機械翻訳の精度を向上するための文章の記載の仕方の工夫としてよく言われる「一文を短くする」、「主語を記載する」等については、出願人が理解しやすい文章につながる側面もあるかと思いますが、先ほど、翻訳を意識し過ぎると日本語がぎこちなくなってしまうというご指摘がありましたように、機械翻訳に最適化された文章が必ずしも出願人の理解しやすい文章であるとは言い切れない場合があるのではと考えています。

次に、特許庁における起案文書に関連した機械翻訳システムの改善についてですが、過去2年間に渡り、進歩



性および記載不備を中心とした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ってきました。

特許庁としては今後も可能な限り機械翻訳の精度向上等に取り組んでAIPNの利用促進を図ることにより、特許審査の国際的なワークシェアリングに貢献していく所存です。

研究開発現場での取り組み等について

潮田：10月4日（日）投稿

守屋所長から機械翻訳に対するユーザー側の期待・要望につきまして、豊富な話題をご提供いただきましたので、それらに関連する研究開発現場での取り組み等についてお話をさせていただきます。

まずどのような日本語に対して精度の良い翻訳ができるかという問題ですが、これは逆に**機械翻訳はどのような日本語が苦手かという観点から考えた方が分かりやすいと思います**。但しその理由は苦手なケースの方が少ないからでは決してなく、苦手なものは極端に苦手だからです。苦手なケースには以下のようなものがあります。

1) 訳が前後の文脈に大きく依存する場合

通常人間が翻訳を行う場合は文章内や段落内での話の流れを理解した上で個々の箇所、たとえば1文なら1文を訳します。時には原文の2文を1文に訳したり、逆に1文を2文に分けて訳したりします。しかし現在の機械翻訳は基本的に文単位での翻訳しかできません。ある文を翻訳している時には、その文の前後でどのような内容が語られているかということを考慮してはしません。従ってたとえば指示代名詞が具体的に何を指しているかなどは分らず、訳に反映させることもできません。指示代名詞についてはまだ読み手の人間が判断できるかも知れませんが、文脈によって訳語が違って来る場合などはやっかいです。たとえば、「…が合成され、その結果、…」という文を英語に機械翻訳する場合を考えてみます。「…」の部分が何かによって「合成され」の部分の訳は“composed”になったり“synthesized”になったり違ってきます。その辺の判断は、前述の訳語選択ルールの活用などにより、今の機械翻訳でも対応可能ですが、もしこの文が「…が合成される。そしてその合成の結果、…」のように2文に分かれていたり、更に2文の間に他の文が入っていたりした場合、2文目の「合成」の訳出が正しく行えるとは限りません。訳語選択以外にも訳が文脈に依存する例はいろいろあります。

たとえば文構造の解釈自体が文脈に依存するような場合もあります。訳語選択については、訳語選択ルール以外にも統計的な処理法などいろいろな検討が行われていますが、文脈処理という大きな問題に対してはまだ現在の機械翻訳は解を見出せてはいません。

2) 長い文の場合

これは説明の必要もないかも知れませんが、機械翻訳は長い文は大変苦手です。それは、文が長くなるほど、文構造の解析（構文解析）が複雑になりかつ精度が低下するからです。一般に構文解析の複雑度は、文の長さの3乗に比例します。文が10倍に長くなれば複雑度は1000倍に増えます。実際には極めて長い文に対しては、適当な切れ目を探して文を分断する等の現実的な処置が施されていますが、原理的な難しさは変わりません。**長文に対する万能薬はありませんが、特許文書の場合抄録文や請求項の文において、定型性の高い文が多く現れるため、その定型性を利用して構文解析の精度を上げようという試みは行われています。**

3) 並列表現が多い場合

複雑な並列表現は文の構造を曖昧にし、構文解析を困難にします。比較的簡単な「AのBとC」と言った表現でも、「Aの(BとC)」と「(AのB)とC」の2通りの解釈が可能です。「私の父と母」や「大統領の側近と財務大臣」のような例を考えると分かりやすいでしょう。ちなみに、よく使われる3つのweb翻訳サイトで「大統領の側近と財務大臣」を訳して見ますと、

サイト (1) : Close to the President and Minister of Finance

サイト (2) : An aide and the Minister of Finance of the President

サイト (3) : President aide and Minister of Finance

のようにかなりバラエティーに富んだ訳文が得られます。サイト (1) は「側近」の訳を間違えてはいるものの並列構造自体は正しく解釈しているように見えますが、サイト (2) は「(AのB)とC」と捕えるべきところを「Aの(BとC)」と誤って解釈しています。サイト (3) は正しく捕えています。

もう少し複雑な例で、並列表現にギャップがある場合はなお解析が難しくなります。たとえば「AをBにXをYに変換する」と言った場合、「AをBに」の後に「変換し」が省略されていると考えられます。再び上記3サイトで「AをBにXをYに変換する」を訳してみると

サイト (1) : A to B to convert X to Y

サイト (2) : I convert X into Y in B in A

サイト (3) : A is converted in B and X is converted into Y

という結果になり、この並列構造（とギャップ）を正しく捉えて訳しているのはサイト（3）だけでした。実際の特許文書の中にはこれらの例よりも遥かに複雑な構造をした並列表現が多く現れるため、並列表現の処理は大変ハードルの高い問題だと言えます。同時に古くから自然言語処理の研究者が着目し、今でも活発に研究されている問題でもあります。

その他にも、主語や述語など文の主要な要素がない場合や、要素の区切りや対応がはっきりしない場合、「てにをは」が欠落している場合、語順が通常の言い回しと違っている場合などが挙げられます。

精度の良い翻訳が得られるような日本語を書こうと考えた場合は、上記のケースの裏返しで、「なるべく短い文で、1文1文意味的に独立させて、複雑な並列表現は避け、並列表現等で省略されている箇所は補い、主語、述語や”てにをは”は省略せず、文の主要な要素の区切りや対応をはっきりさせ、なるべく自然な語順で書く」のが望ましいということになろうかと思えます。なお、商用機械翻訳ソフトの中には、一まとめにして訳して欲しいと思う箇所は『[大統領の側近]と財務大臣』のように、[]などの記号で囲うことでその通りに訳してくれるものもありますので、長文や構造の複雑な文を訳すときには便利です。

次に、「少なくとも人にとって明晰な日本語文の直訳機能を向上させる」べき、というご指摘についてですが、これに似たご意見ご要望はこれまでも多くのユーザーからいただいております。「簡単な文だけで良いから完璧に訳して欲しい」と言ったご意見などです。もちろん機械翻訳の研究開発現場ではそう言ったご意見は十分認識されているわけですが、**人間にとって簡単な文と機械にとって簡単な文が必ずしも一致しないのが難しいところ**です。機械にとって難しい文の特徴は上に挙げた通りですが、確実に訳せるような易しい文とは何かとなりますと、実は機械翻訳ソフトごとにそれぞれ違ってきます。共通して言える特徴を強いて挙げるとすると、曖昧性のない文です。言語とは本質的に曖昧性を内包しているものですが、ここでいう曖昧性のない文とは、単語・複合語、フレーズなどの語句の切れ目や構文の解釈などが一意に定まり、それぞれの語句の訳語が一意に定まるような文です。但しこれは機械にとって一意に定まるという意味なので、やはり人間にとってはストレートに役立つ情報ではありません。たとえば、あるフレーズが名詞句的にも副詞句的にも用いられ得る場合そこに曖昧性が生じますが、人間がその曖昧性に気づくのは難しい場合がほとんどです。具体例でお話したいと思います、（機械翻訳ソフト自体の改良の課題は）「人間にとっての明晰な和文の直訳機能の不十分さです。」と指摘された大野副本部長が使われた例をもとに考えてみましょう。大野副本部長のご発言から引用します。

『「画像データが保存部 207 にあるかどうかに基づいてこの判断は行われる」という、人間にとって明晰な和文を某翻訳ソフトで機械翻訳した結果、”The image data is in preservation part 207 or this judgment is done based on” と訳されてしまいます。（希望英文：”The determination is performed based on whether image data is stored in storage 207.）』

大野副本部長が使われた「某翻訳ソフト」のパフォーマンスを私が再現することは出来ませんので、今別の機械翻訳ソフトで同じ原文を訳して見ますと、

“This judgment is done whether the image data to be in preservation part 207 based on.”

という訳が得られました。程度の差こそあれやはり全く意味の通じない文になっています。しかし、ここでは2つの機械翻訳ソフトが共通して直面していると思われる、機械にとっての難しさ（曖昧性）が見て取れます。それは、「画像データが保存部 207 にあるかどうか」が名詞句なのか副詞句なのかという判断の難しさです。人間にとっては、「あるかどうかに基づいて」と言えば「あるかどうかという点に基づいて」という意味であり、ここでは「あるかどうか」は「あるかどうかということ（点）」という名詞句として用いられていることは明白です。しかし機械にはこの判断が難しいのです。何故かと言いますと、「～かどうか」という表現は通常の名詞句とは異なり、副詞句的に用いられることもあるからです。たとえば「真の目的がそこにあったのかどうか多くの米国民は懐疑的だ」と言った場合、「あったのかどうか」は名詞句としては用いられていません。名詞句の後に何の助詞も伴わずに「多くの米国民は懐疑的だ」という表現が続くことはないからです。

そこで、元の文の「あるかどうか」に相当する部分が名詞句であることを明確にするために、原文を

画像データが保存部 207 にあるかどうかという点に基づいてこの判断は行われる」

と書き換えてみますと、機械翻訳の訳文は

“This judgment is done based on the point whether the image data is in preservation part 207”

となり、構文が正しく捕えられていることが分ります。

あるいは主語が「この判断」であることを明確にするために、原文を

「この判断は画像データが保存部 207 にあるかどうかに基づく」

と書き換えますと、今度は「かどうか」が名詞句であることを正しく認識して

“This judgment is based on whether the image data is in preservation part 207”

という訳文が得られます。

このように、人間にとってごく単純で明晰な文であっても、機械にとって何らかの曖昧性が生じる場合には、思うような訳文が得られないことが間々あります。また

曖昧性を持った同一の表現でも、短い文やその他の曖昧性の少ないような文の中では正しく訳されても、別の文中に現れると正しく訳されない、といった現象も起こるため、正しく訳せる文といったものを一般的に語るのは非常に難しいのが現状です。開発の現場では、このようなうまく訳せない事例を1つ1つ検証し、それぞれについて曖昧性を解消する手立てを考案して改良するというステップが繰り返されています。

技術分野による用語の訳し分けに関しましては、前述のようなコーパスを用いた訳語選択ルールや統計的な訳語選択モデルを、分野ごとに構築していくのが有効な解決手段だと思います。時代の変化による技術用語の変遷への柔軟な対応は、コーパスを適宜アップデートすることにより実現できると考えられます。

権利解釈等で不利益にならない訳語選択を行うためには、技術用語の意味概念の領域と技術用語間の意味的な関係を正しく捉えるためのオントロジーだけでなく、オントロジーと入力文の構文情報や意味構造を組み合わせて適切な訳語を選択するためのメカニズムを構築する必要がありますが、これは現状の機械翻訳や自然言語処理のレベルでは至難の業と言えるでしょう。また権利解釈を左右するのは技術用語だけではなく、一般的な表現も重要な役割を担います。たとえば「基板上に形成する」と言った場合、“on the substrate”と訳せば基盤に接触していることが条件になりますが、“over the substrate”と訳せば離れて上という意味も含むなど、権利解釈上重要な差異が生じ得ます。このような判断を正確に行うには、単に一文一文を正確に訳すという範囲を超えた、技術上の高次の判断が必要であり、機械翻訳にその任を託すことには無理があるでしょう。むしろ



る現実的なアプローチは、高次の判断は人間に任せ、機械は、人間が見落としがちな様々な曖昧性を自ら解消するのではなくユーザーに注意を喚起して、可能なオプションをユーザーに提示するという役に徹することだと思います。

最後に非英語圏の言語の中で代表格の中国語の機械翻訳についてですが、研究の歴史が英語の機械翻訳に比べて浅いこと、および中国語という言語自体の持つ特殊性により、日中・中日機械翻訳の精度はまだ日英・英日機械翻訳のレベルには達していません。中国語の特殊性というのは、中国語だけが特殊だということではありませんが、これまで機械翻訳の研究開発において中心的に扱われてきた英語や日本語に比べて、中国語は文法構造と意味構造が密接に絡んでいるという点で特徴的です。別の言い方をしますと、特に英語などは、意味が通じるか通じないかということとは比較的独立に、ある文が文法的に正しいか正しくないかという判定が行えますが、中国語では両者が密接に関わっていて、意味的な解釈の妥当性を抜きにして文法的な正しさを判断することはできません。ちなみに日本語はその中間に位置していると思います。従って、中国語の構文解析を正確に行おうとした場合には意味処理の要素をも取り込む必要があり、難易度が増します。

しかしながら、中国語の特許翻訳に関しましては、抄録文や請求項において、日本語の特許と同等かあるいはそれ以上の定型性が見られる傾向にあり、将来的には上述のような定型利用機械翻訳の枠組みによる大幅な翻訳精度の向上も可能ではないかと期待されます。

特許のトータルな系でのシステム構成、明晰な日本語の定義などが重要

辻井：10月4日（日）投稿

機械翻訳というのは、非常に単純化してしまうと、翻訳元の言語の単語を翻訳先の言語の単語に置き換えて、単語の並び方も、翻訳先の言語の並び方に変える、という技術です。もちろん、一方の言語であからさまに表現されていることがもう一方の言語では表現されない、たとえば、会話の文で英語での主語がIやyouなどの場合、日本語文では省略されて消えてしまう場合などがありますから、これほど単純ではないですけど。

単語の順序をかえる部分は、文が短いほど、すなわち、文中の単語の数が少ないほど、それだけ可能な並べ替え方の数自体が少なくなりますから、正解の翻訳を出しやすくなることは確実です。また、人間は無意識のうちに文法を使って単語間の構造を意識して読みますから、文法的に正しい文であれば、多少長くても理解できます。ところが、長い文で単語の並べ方が誤っていると、理解が非常に難しいということもあります。一般的

に、短い文のほうが人間にとって理解しやすいですから、機械翻訳の出力文のように誤りがある文では、特に、短いほうが理解できやすいということになるでしょう。したがって、機械翻訳にかけられる文は短いほうがいい、というのは一般的に成り立つと思います。

もう一つの、翻訳元の単語を翻訳先の単語に置き換える過程は、もとの単語が一般語で意味の幅が広いほど、文脈に応じて相手言語での単語を選択しなければならず、機械にとっては難しくなります。

これは、人間の場合とは少し違いますね。非母国語の人にとっては、一般語で日常的によくつかわれる単語を使うほうがたくさんの単語を記憶しなくてもよいので、Simplified Englishのような非母国語話者のための言語では使える単語を制限するのが普通です。でもこれは、上で述べたように、機械翻訳にとっては負担です。定義が明確で意味の幅が狭い単語のほうが、翻訳辞書も容易に作れるので、適切な単語を文脈に応じて選ぶ必要がないだけ簡単になります。ここでは、機械翻訳のための明晰な特許文と人間が理解しやすい特許文の書き方がずれてくる原因となるでしょう。

このような人間と機械ずれは、他のところでも出てくる可能性はあるでしょう。たとえば、日本語は主語や目的語が省略されて、英語に翻訳するときに困ることがあります。機械翻訳の立場からは、省略を許さないという制限が考えられます。ただ、このように制限を付けると、不自然な、ごつごつとしてわかりにくい日本語になってしまうでしょう。省略というのは、後ろの文で主語が省略されることで、2つの文の間の関係がより緊密になり（テキストの結束性といいます）、複数の文からできたテキストのテキストとしての自然さを増すという重要な機能を持っているのです。省略すべき個所で省略しないと、日本語としてはわかりにくい、ということになります。

個人的には、わかりやすい日本語で特許を書くというとき、いろいろな条件を付けた文法で日本語のテキストを書くというイメージは、機械・人間にとって共通の規則（なるべく、短い文でかく）を規定しておいて、それ以外の機械と人間との間に「理解しやすさ」に差がある部分では、人間に見せる表面上の文と、機械のために表面からは隠れていて注釈として付けられている情報といった二重性を持たせる必要があると思います。

このようにいうと、表面上みえない情報をテキストに付記するという、非常に厄介な作業のように思えますが、最近のタグ付けツールの展開などを考えるとうまいインターフェースを設けることで、あまりコストがかからずにできると思います。

それ以外の、機械翻訳が理解しやすい日本語の条件というのは、対象の機械翻訳がどういう作りで作られたものかに依存するので、一概には言えない部分があります。ただ、AAMT/Japio 研究会で特許文に構造をふるという作業をしていると、人間でも、その分野のことを

かなりよく理解していないと正しい構造が決められない、という状況があります。機械翻訳に実際に役に立つかどうかは別にして、人間による解釈に曖昧さをできるだけ残さない文を書く、という特許の目的からするとどうかなという文が数多く出てきます。どういところで**解析者が解析に迷うかという観点で、特許文への構造づけの作業を行えば、結果として「曖昧さの生じない」特許文として、どのような規則を設けるかが明らかになり、それが明晰な日本語というものの定義につながる**と考えています。

もうひとつ、英語以外の言語への翻訳ということに関連して、今、日本語—中国語機械翻訳システムを科学振興調整費で研究を進めていますが、英語以外の言葉への機械翻訳システム開発の困難さを身にしみて感じています。出力された中国語が理解できないから、中国人研究者にチェックしてもらわなければならない、チェックの結果を実際のシステム開発に反映させるところで問題がでる。単に出力された中国語が理解できる、できないだけでは、翻訳過程のどこで失敗したかがわからない、したがって、システムを改善するためのフィードバックができないのです。このようなことは、機械翻訳を実際に運用する場合にも生じると思います。プロジェクトでは、日本語・中国語の専門用語の辞書を英語を介在させて作っているのですが、翻訳システムの運用時も、英語を介在させるしかないのでは、と考えています。つまり、**日本語から中国の翻訳をつくる際に、同時に英語の翻訳も作って、英語の翻訳が正しいことから、中国の訳も正しいというある程度の保証を得る、といったことを考える**を得ないと思っています。

ただ、この場合には、日本語から中国語への翻訳過程と英語への翻訳過程に平行した部分が多くあって、翻訳の英語が正しい場合には中国語も正しいと（ある程度）



判断できるようなシステムの作りになっている必要があるかと思います。この点では、現在、米国中心で研究されている統計的機械翻訳は、基本的に2言語間の翻訳モデルになっていて、この保証がないので、特許のような翻訳の質をある程度保証しないとだめな応用ではまずいのでは、と思っています。

いずれにしても、**機械翻訳の基本的な技術が成熟してきたこの時期に、特許を書く人まで含めたトータルな系で、コストを最小にするシステム構成、明晰な日本語の定義などが重要になってきている**と思っています。

ブレークスルーは、言語処理の専門家と知財の専門家との協働から！

岩田：10月5日（月）投稿

外国出願の明細書作成をなかなか効率化することができない日本企業の現状と課題にフォーカスして、少し詳しく述べさせていただきます。

1回目の投稿では、外国出願の明細書を作成する際に機械翻訳システムはほとんど使っていないと申し上げたのですが、では、なぜ使わないのかについて、少し詳しくお話いたします。

まず、日本語特許明細書を翻訳しやすいように書き直した翻訳用日本語原稿、これを機械翻訳システムにかけても、上手く翻訳してくれないことが少なからずあります。**特許の専門家は、言語処理の専門家ではないため、機械翻訳システムが日本語の構文を捉える方法や規則が解からず、翻訳用日本語原稿をどのように修正したら機械翻訳システムが上手く対応してくれるのか、見当が付きません。**

また、機械翻訳の結果が英語としては間違っていないとしても、特許明細書での英語表現としてより好ましい表現があることがよくあります。しかし、その表現を機械翻訳システムに上手く憶えさせる手だてが見つかりません。

さらに、日本企業は、知財部員を米国等の現地代理人事務所へ定常的に派遣する場合がありますが、派遣された知財部員は、現地の弁護士や弁理士から指導を受け、明細書作成に関する最新のコンセプトを持ち帰ります。日本企業は、それを消化吸収して、各社それぞれに、独自の外国明細書作成ノウハウを持ち合わせています。そのノウハウにはさらに、他社とのライセンス交渉における苦い経験からのフィードバックも含まれています。しかし、**現在の機械翻訳システムに、そのような各社それぞれのノウハウを反映させることもなかなか難しい状況**です。

結局、外国出願の明細書を作成するにあたり、機械翻訳のユーザーとして現状でできることは、機械翻訳シ



テムが所望の英語文章に変換してくれるような日本語原文を、手探りで探し出すことしかないわけです。それも多くは上手くいかず徒労に終わりますから、結局、話せば解かってくれる人手の翻訳者に頼り、機械翻訳から離れていく、これが現状だと思っています。

当社は機械翻訳のベンダーでもありユーザーでもあるわけですが、ユーザーの立場からすると、**言語処理の専門家からなるベンダー側と知財の専門家からなるユーザー側とが乖離**していて、ベンダー側から新しい機械翻訳システムが発表されると、ユーザー側は試しに使ってみるが、その中身はユーザーにとってブラックボックスですから、上に述べたような問題点を解決したりフィードバックしたりする手だてもなく、諦めて、人手による翻訳へ戻っていく、この繰り返しになっているのではないかと思います。ベンダー側の長年にわたる研究開発の成果を、知財の分野においては活かすことができていない、残念な状況です。

ベンダー側とユーザー側とが互いに歩み寄ることができる新しい施策を打っていくことで、この状況を変えていく必要があると思います。

以上のような状況を踏まえて、外国出願の明細書作成を効率的なものにしていく方向性について、考えるところを述べさせていただきます。

まずは、「特許版産業日本語」と「特許明細書ライティングマニュアル」の取り組みですが、これは、私の知る限り、機械翻訳のベンダー・研究者と知財関係ユーザーとの**はじめての本格的な協働の取り組み**です。

最初は欲張らないで、ユーザー側が日本語明細書を翻訳しやすいように書き直す翻訳用日本語原稿、これを知財専門家であるユーザーが意図するとおりの正確さで

機械翻訳することができるような、日本語仕様の策定を目指すのがいいと思います。両者が話し合う中で、ユーザー側は機械翻訳システムの得手不得手を理解し、例えば日本語の文章そのもので構文や係り受けを正確に示すことが難しいのならば、ユーザー側が譲歩して、翻訳用日本語原稿の中に構文や係り受けを指示する記号を導入することなども、あっていいと思います。

外国出願明細書作成の効率化の第一歩として、「特許版産業日本語」と「特許明細書ライティングマニュアル」の制定を目指す現在の取り組みは、ぜひ成功させていきたいと考えています。

つぎに、**翻訳用日本語原稿をオーサリングする仕掛けが必要と考えます。**

「特許版産業日本語」と「特許明細書ライティングマニュアル」が完璧であったとしても、ユーザーが実務でその仕様を厳格に守ることは、却って非効率かもしれません。また、それらの仕様にすべての事例に完全に対応できるだけの網羅性を持たせるのも、現実的ではありません。

そこで、ユーザーが産業日本語仕様に則して作成した個々の翻訳用日本語原稿を、インタラクティブに添削して、翻訳で問題が出そうな箇所を指摘し、修正を要求し、修正案を提示してくれるような仕掛け、このような**オーサリング・システムをイメージして作り込むことが必要だと考えています。**

知財ユーザーとしても、特許版産業日本語をマスターしたうえで、このようなシステムを使いこなすことで、自分の日本語の癖を認識して修正しながら、正確に翻訳されるために必要な日本語の書き方を習得することができるものと思います。

そして、最後に機械翻訳システムですが、特許版産業日本語仕様に則して書かれ、さらにオーサリング・システムによって修正された日本語文章であっても、機械翻訳システムにとってはいまだに翻訳しにくいということが、あるのではないかと思います。

この点、現状の機械翻訳システムは、ボタンを押したら頑張って一発で翻訳結果まで出してくれるのですが、多くの場合、結果は満足できるものではありません。このいわば**一発回答方式の頑張り**は、**必要ないのではないかと考えています。**翻訳しにくいならば、ユーザーに対してその点を指摘し、日本語の修正を要求してくればいいのかと思います。

さらに、先に述べたような、各社各様の外国明細書作成のノウハウ、これを、翻訳の時点で、機械翻訳システムが反映してくれると有難いと思います。ユーザー毎のカスタマイズ仕様を機械翻訳システムが受け入れてくれることも必要なのではないかと考えます。

以上、ユーザーの立場から考えるところを述べさせて

いただきました。最後に、ブレイクスルーのキーは、言語処理の専門家と知財の専門家とが、直接に、あるいは特許版産業日本語仕様や実際の機械翻訳システムを介して協働を行うことである、ということを通り越して強調させていただいて、発言を終わります。

“青木ヶ原樹海”、“翻訳の使命”、“80点主義”

中山：10月6日（火）投稿

三題話的になって恐縮ですが、今回のテーマ“わかりやすい明細書や機械翻訳しやすい明細書を書くにはどうすれば良いか、さらには、不可侵領域とされる（!?）特許請求の範囲の記載を明晰にするにはどのような工夫が考えられるか”が示されましたときは、咄嗟に“青木ヶ原樹海”、“翻訳の使命”、“80点主義”の言葉を頭をよぎりました。

1. 富士山麓の青木ヶ原樹海との関連

“不可侵領域とされる（!?）特許請求の範囲の記載を明晰にする話に一旦踏み入れたら出るにでられないかもしれない”という思いです。

この領域は、単に翻訳技術だけを論じても片手落ちという気がします。

“日本オリジナルの発明”について“日本人”が“日本語で書いたもの”を“日本の裁判官が解釈する”ケース、あるいは、“米国オリジナル発明”を“米国の人が”“英語で書いたもの”を“米国の裁判官が解釈する”ケース、いずれも母国語だから全く問題は生じていないという構図にはなっていません。この話は、翻訳の問題とは全く関係ありませんが、言葉で書き表した際に生じ



る問題、すなわち翻訳以前の問題もこの領域にはあることを前提で考えてみる必要があります。

発明した人の思いを明細書の書き手が書き表した時点で何らかの“差 (1)”が生じている、それを解釈する人との間で“差 (2)”がでてくる。前で生じた差 (1)の差が縮まる方向に動けばいいのですが、差 (2)ではさらにその差が増幅される例が多いと思います。

少々脱線しますが、現役時代、個人発明者の方から「自分の発明を勝手に実施している。けしからん」とのクレームを度々受けました。当該技術分野の知財担当者を伴って会議の席に出て、「どこに発明者の方が説明する発明が記載されているのか分からない」との回答をすると、さらにお叱りがくる。「自分のアイデアを弁理士さんに書いてもらっているの間違いはないはず、答えをはぐらかすものではない」と。弁理士さんの書き方が悪かったとは思えません。発明者の方のアイデアが書き手側に正確に伝わらず、書き手側では、正確に表現することが難しかったのかもしれない。このようなやり取りは、企業同士でも生じます。

このように翻訳の問題の前に発明者⇒書き手の間に差が発生しています。さらにその先には翻訳文が出願される国の裁判所における解釈が待ち受けています。“翻訳”はこの一連の連鎖の途中に位置するものであり、翻訳の問題をいくら詰めても、この一連の問題を一気に解決する術にはなりません。

つまり、“しかれば、”翻訳の使命“はどの辺りにあるのでしょうか”という疑問を提示したかったので、回りくどい話をしました。

2. 翻訳の使命

“パーフェクトな翻訳”はありうるのでしょうか。また、ここでいうパーフェクトとは何でしょう。

先ほどは、発明者、書き手という登場人物だったのですが、翻訳の局面になると翻訳者という新たな人物が登場してきます。翻訳者は、書いてあることに忠実に翻訳しようとしていると思います。その姿勢は、全く疑う余地はありません。しかしながら、“幾多の翻訳のありようから”ひとつを選んだときに、翻訳者の無意識な意識が入り込み、“書いてある日本語”との差が生まれてきます。

またまた、脱線して恐縮ですが、中国で通訳を付けて会議する際、自分が話した時間よりも2倍3倍掛かって通訳されている状況がよくあります。その逆もしかりで、中国の方が話された内容について通訳の方は相手方(話し手)と長々と話しをされていて(通訳の方が質問しているケースがおおいのですが)、我々へは、“話し手はこういつている”となんと簡単に紹介されるか。なんと通訳の意志がかなり入った会話になっていることか。書き物の世界(書き表された明細書とその翻訳)ではこういうことは起きていないといい切ることができるでしょうか。

機械翻訳は、人間の意志が入らないので、より正確だということもできません。よしんば、機械翻訳は、人間の意志を除くのでよりましたとしましても、パーフェクト領域には到達できないのではないのでしょうか。

英語に翻訳されたものが米国の裁判所で解釈される。米国での裁判でいい結果がもたらされなかった場合、翻訳の依頼人からみればオリジナルとの齟齬が生じていたため(翻訳が悪かったため)ではないかと疑心暗鬼になる。(翻訳者からすれば、元が悪くて翻訳者の方が苦勞して翻訳したのにとの抗弁が聞こえそうですが。)

したがって、**裁判で優位な立場をキープできたとき、出願人(翻訳依頼人)は、一連の出願プロセスに満足を覚え、はじめて翻訳についても“満足”との評価をするわけです。これから言えるのは、翻訳の出来具合は、相対的に評価されるものであり、パーフェクトな翻訳はありえないということです。**

3. 80点主義

会社に入社したばかりで仕事のアウトプットが遅かった頃、「パーフェクトを求めるものじゃないよ。パーフェクトはありえないのだから、80点で切り上げる」と言われたものです。「ぐずぐずして出さぬより、早く形を見せろ。手を入れてやるから」と80点主義を叩き込まれました。確かに自分で100点と思っても、他の人の手が入るわけです。翻訳の世界でも同じではないのでしょうか。

複数の登場人物がいて、個々のプレーヤーが100点と思っても、それぞれの間には差が発生している。これら差を埋めることは、100点を狙った機械翻訳を導入だけでは(たとえそれが完成したとしても)解決できません。

したがって、**クレーム部分の翻訳問題は、“80点主義”で考えてもいいのではないのでしょうか。すなわち、発明者の意思(“自分のアイデアはこうだ。それを何とか立派な権利にしたい”)との切実な思いを正確に(箇条書き的に)書き表し、それを翻訳する。それをクレームの原型とする。日本の制度に合わせ、日本語でこねられた(練られたというべきかもしれませんが)クレームを機械で翻訳することまでは求めない。**

このレベルであれば、発明の詳細な説明をより正確に翻訳することを目的に開発された機械翻訳ソフトウェア(今、改良のために努力されているソフトウェアの完成形)で十分なのです。米国出願のクレームは、米国特許弁理士にドラフティングしてもら(クレーム作りは任せる)ので、米国の特許弁理士との正確な意思疎通が図れば良とする。

ここでの機械翻訳の効果は、翻訳の専門家に任せなくても、万人が、思い通りの米国出願クレームを作成する道が開けることです。安く上がるかは定かではありません(米国の専門家の仕事の値付け次第ですから)。クレームまで機械翻訳で対応できれば、相当なコスト低減にな

翻訳にまつわる問題点

- 意識なのか、書いてあることを翻訳したのかの判断
⇒この差と境界線、誰が判断するのか

* 言語体系が異なるので完全一致の翻訳はありえない
- 誤訳なのか、書いてあることを翻訳（採択）したのかの判断
⇒裁判では、この差が issue になる

* 権利解釈において、単なる字句の訂正とそうでないのは大違い
- パーフェクトな翻訳とは、ありうるのか
⇒それを判断する者と基準

* 絶対的基準はありえず、相対的なもの（裁判で勝てればそれでいい、負ければ翻訳技術で負けたのではと疑う）

機械翻訳に期待するもの

- パーフェクトな翻訳をもとめているのか（現行の日本語出願明細書に何らかの手を加えることなく、人の手による翻訳を超えるものを目指す）
* クレーム翻訳は射程にないことを前提として
- 機械翻訳の精度を上げることは何に繋がるのか
⇒素人が玄人の翻訳技術を入手する
* 辞書の充実により、より適正な技術用語の選択が可能になることを前提

登場人物とその人たちの気持ちのすれ違い

- 発明者と書き手の関係：[発明者] としては自分の発明が過不足なく書けているのかとの心配
[書き手] としては、発明者の考えているものを全部引っ張り出したのに評価されないという不満
- 書き手と翻訳者：日本語明細書の [書き手] としては、翻訳によって発明の内容が変わってはいけないとの心配
[翻訳者] としては、元の日本語がきちんとしていないので最大限の努力をしてもここ止まりという気持ち
- 外国代理人と出願依頼者：[出願人] としては、本当にいい出願をしてくれたのか、いい権利を取得しようとしてくれたのかとの心配
[外国出願代理人] としては素材が悪ければ努力にも限度があるという気持ち
- 裁判官と権利者：[権利者] としては、発明の本質を理解してくれないとの嘆き
[裁判官] としては、表現されていない内容の権利は評価しようがないという思い

登場人物が大胆になれない事情

- オリジナルと翻訳の差と優先権を使える限度の問題
⇒内容が変わっていれば優先権が認められないが、その解釈と基準
* 詳細な説明でカバーされる範囲の見極め
- 翻訳でオリジナルより分かりやすい表現を用いたとき
⇒オリジナルとの違いによって生じた差が、悪い方向に向かわないか
向かった場合、誰が責任をとるのか
- クレームを平易な文で表した場合の問題点（翻訳が容易になるという利点が生まれる）
⇒裁判でのクレーム解釈に影響を及ぼすことはないかとの疑心暗鬼
- 権利解釈に影響を及ぼすか否かの実験は不可能
⇒裁判で判決を得るまでは分からない。裁判に持って行く案件は会社として重要なものであるため、実験というリスクは犯せない。

るのでしょうが、上述のような理由でそこを解決しても他の問題があるので、多くを期待しなくてもよさそうに思います。前回述べたところとの食い違いがあればご容赦をお願いします。

日本語のかなり練られたクレームと翻訳文（英文明細書）との一致が果たせてないので「優先権主張が大丈夫か」とのご指摘もありましょう。言語体系が異なるので完全一致はありえないという前提で考えれば、発明の詳細な説明でサポートされていればその心配はいらぬとの見方もあるでしょう。もしそうでなければ、これを認めるべきとの世論形成に向かう必要があるように思います。

ここの部分が、いただいたテーマにある“ある種チャレンジングな取り組みの可能性や問題点”の部分になると思います。それぞれの国の国語に合わせてベストの表現を採用したさいの齟齬、発明の表現方法が少し変わってくることをネガティブに考えず、翻訳された第二国出願にオリジナル出願と異なった権利要求部分は、発明の詳細な説明の範囲にあることを出願人に疎明させる。それはそれで良とするスキーム作りができないものでしょうか。

本件に関して考えた事項を参考までに記載します。

おわりに 機械翻訳というチャレンジングなテーマを巡り、多くの示唆に富んだコメントやアドバイスに感謝

進行 守屋：10月9日（金）投稿

皆様、お忙しい中での二巡目のご投稿、誠にありがとうございました。

それでは、この座談会の締めくくりとして、皆様から



いただきましたご意見やご指摘を参考に、私なりに思うところも加え、論点を整理しました。

（1）特許分野における機械翻訳の活用の現状

特許行政の現場では、我が国はもちろん、各国特許庁においても、機械翻訳システムを実際に活用し、特許審査の効率化やユーザーサービスの向上に役立てています。特に、「特許審査ハイウェイ」においては、機械翻訳が出願人の翻訳コストの負担を軽減しています。このような、特許行政の現場での、積極的な機械翻訳活用の動きに、機械翻訳の研究開発に携わる者は、自信を持つと同時に己を鼓舞することも忘れてはならないと思います。その一方で、特許制度や特許情報のユーザーの皆様からは、機械翻訳は、海外特許調査の一部や英語特許文献の概要把握のために活用されているものの、海外への特許出願明細書の作成現場では、部分的な適用の試みや翻訳者の翻訳支援といった活用レベルであること、そして、非英語圏の特許文献調査や概要把握への活用については、行政、ユーザーの双方の側から、機械翻訳の精度は実用の域に達していないことが指摘されています。

（2）今後の機械翻訳システムの研究開発において留意すべき点

①精度の良い機械翻訳結果が得られる日本語文とは？

一般論としては、「なるべく短い文で、1文1文意味的に独立させて、複雑な並列表現は避け、並列表現等で省略されている箇所は補い、主語、述語や”てにをは”は省略せず、文の主要な要素の区切りや対応をはっきりさせ、なるべく自然な語順で書かれた文」であるとのこと回答をいただきました。

産業日本語の検討では、この回答を更に分析し、より具体的かつ詳細に整理する必要があります。

なお、「文の長さ」は、文の複雑さを示す一つのわかりやすい指標と言えそうです。

②機械翻訳に知財ユーザーが望む機能～インタラクティブな機械翻訳システムの実現～

特許分野のユーザーは、何とか機械翻訳システムを使いこなして、翻訳の効率を上げ、そのコストを下げたいと考えています。しかし、現状の機械翻訳システムでは、システムが日本語の構文を捉える方法や規則が解からず、日本語原稿をどのように修正したら良いのか見当が付きません。

特許分野のユーザーが求めていることは、入力された日本語文に対して、まず、機械翻訳システムが把握（解析）した文の構造や曖昧な箇所（複数通りの解釈が可能な箇所）を提示し、ユーザーに注意や修正を促すこと。そして、再度のユーザーからの翻訳指示により、翻訳結果を表示する。こういったインタラクティブな機械翻訳システムを早急に実現することかと思えます。

こうすることで、「テキストを書いた人の意図」を機械に入力することができますし、「一般用語の訳し分け

の難しさ」についても、機械翻訳システムが、訳語の候補を表示し、ユーザーに選ばせる仕組みを取り入れることで、ある程度解決できるのではないかと思います。

③中国語等の非英語圏の言語に対する機械翻訳の精度の向上

中国語を対象とした機械翻訳は、まだまだ解決すべき課題が多いようです。言うまでもなく、中国市場は日本の産業界にとっても重要な市場であり、特許等の出願件数も爆発的に増えています。今後、中国への特許等の出願や、中国における他社の出願動向、特許権の存在状況等の調査の重要性が増すことは明らかですので、是非とも研究開発を頑張ってくださいと思います。

④機械翻訳システムの進歩をより加速するために

技術の進歩は、その技術がユーザーが使うことで、新たな要求が生れ、その要求にこたえるために更なる技術開発がなされるという連鎖の上に成り立ちます。機械翻訳システムの進歩についても、開発されたシステムを大いに使用しその効果をアピールすべきですし、活用法を工夫して使用するなどして、開発者もユーザーも機械翻訳を市場に弛まず浸透させ続けることが、進歩を加速することになるでしょう。その際に、最初から100点満点を望むのではなく、80点レベルでも良しとして、段階を踏んで確実に進歩させていくことが必要とのご指摘は、まさにそのとおりだと思います。

(3) 特許版産業日本語の開発や特許明細書ライティングマニュアルの作成に際して

①なぜ産業日本語が必要なのか？

「発明者、書き手、翻訳者、外国裁判官といった登場人物の間で、伝達される情報の差分の問題がある以上、この一連の連鎖の途中に位置する“翻訳”の問題をいくらか詰めてもしょうがない」とのご指摘は、まさに、産業日本語のような書き言葉の必要性を提起するものと考えます。目指すべき特許版産業日本語は、発明者が考えた“発明”を日本語文で正確に記述し、正確に明細書の読み手に対して伝え、かつ機械処理にも適するという機能を持った術でなければなりません。

②特許版産業日本語の仕様をつくるにあたり留意すべきこと

まずは、特許明細書の技術説明部分の文章を対象に、作者の意図を正確に伝え、特別な背景知識を持たない人でも理解でき、その解釈が一意に定まり、かつ機械翻訳にも適した複雑でない文を工夫し、次に、そのような文への書き換えを手順だてて行う手法を見出すといった段階的な検討が必要と考えます。このような検討を重ねた結果、明晰な日本語文章とその作文手法が定式化され、オーサリングシステムの具体的な設計（詳細設計）に入れると思います。

③将来の特許版産業日本語オーサリングシステムの開発に向けて

最終的な目標は、「ユーザーが作成した明細書に対

して、翻訳で問題が出そうな箇所を指摘し、修正を要求し、修正案を提示するといった、インタラクティブに行う仕掛け」を持ったオーサリングシステムの構築です。また、そのようなオーサリングシステムを支える仕組みとして、業界で共有する共通辞書を整備し、そこには、日本語に関する知識の他、特許特有の言い回しや表現パターンとその翻訳パターンも収録し、産業日本語への変換とその翻訳を一度に行える仕組みを検討しなければならないと思います。

オーサリングシステムへの第一歩として、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、用語集等をデータベース化し、オンラインで参照できる環境を提供するとともに、参照したユーザー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たいと思います。こうすることで、ライティングマニュアルの利便性や運用性の向上が期待されるとともに、オーサリングシステムの詳細設計に必要な情報を効率よく集め、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効率よく蓄積できるのではと期待しています。

④特許版産業日本語委員会の活動への期待

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の皆さんが集い、協調し協力しながら作業を進めている特許版産業日本語委員会の活動に対し、皆様からの大きな期待が寄せられていることを改めて認識しました。相手の専門領域をある程度理解し、お互いに歩み寄ることが、よりよい協調・協力体制を作り、アウトプットの質と量をぐっと高める秘訣であると考えます。このような協調・協力体制を作る役目を担えるように、Japio 特許情報研究所は努めたいと考えます。

機械翻訳というチャレンジングなテーマを巡っての今回の座談会におきまして、踏まえるべきご指摘やご意見、示唆に富んだコメントやアドバイスを多数いただきました。ありがとうございました。最後になりましたが、本座談会を終了するにあたり、ご参加いただきました皆様に心から御礼申し上げますとともに、引き続き、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)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。