

弁理士と英語

会員 遠藤 朱砂



要 約

知的財産は、その価値が高ければ高いほどグローバルな保護が必要であり、かつ、企業の事業展開に伴い、世界的な権利取得・維持・行使が活発になっています。各国の知財保護は各国独自の法律で定めた公用語で行われますが、英語が世界の共通言語としての地位を高めている現状では、弁理士業務における出願・審査段階の書類翻訳、通信文、プレゼンなどで英語の重要性が更に高まっています。最近の機械翻訳の精度の飛躍的上昇に伴い、沢山の団体が様々な観点から機械翻訳の活用を提案しています。このような昨今の情勢を踏まえて、弁理士業務における英語の重要性をまとめると共に、NIPTA 検定制度を紹介し、機械翻訳の活用と昨今の人工知能 AI ブームについて説明します。

目次

1. はじめに
2. 知財業界における翻訳の問題
3. 特許関係書類の翻訳
 - (i) 翻訳に求められる要件
 - (ii) NIPTA (NPO 日本知的財産翻訳協会)
 - (iii) 機械翻訳の活用法
4. 意匠・商標関係書類の翻訳
5. 英語での発表
6. 人工知能について
7. まとめ

ローバル傾向は明らかです。⁽¹⁾

世界のいずれかの国で知財の保護を求める場合、各国独自に定められた公用語で手続きを行うのが原則です。しかし、主に英語圏で開発されてきた IT 情報技術の発達により、世界共通言語としての英語の地位は向上しています。従って、日本国での内外・外内案件の処理は、現在かなりの割合は英語で行われています。すなわち、出願・審査段階の書類、通信文、プレゼン資料などを、和英・英和翻訳する必要性が更に高まっています。一方、最近の機械翻訳の精度の飛躍的上昇に伴い、様々な観点から機械翻訳の活用が検討・提案されています。しかし、機械翻訳には問題点も多いため、その利用には注意が必要です。

1. はじめに

知的財産は、その価値が高ければ高いほど世界各国での重要度も高く、グローバルな知財確保の必要性がより高くなります。従来行われてきた、自国製造して海外へ輸出販売する事業モデルだけでなく、本社とは異なる国で研究開発や製造を行い、別の国へ輸出入して、世界中へ販売する、いわゆる事業のグローバル化が進んでいます。そのため、世界中で知財に関する権利を取得・維持・行使する必要がますます高まっています。特許庁の発行した特許行政年次報告書 2017 年度版によると、世界の特許登録件数中の非居住者による登録は、この 10 年間で約 1.6 倍に増加して 2015 年には全体の 4 割弱を占め、2016 年の日本からの PCT 国際出願件数は、2012 年から約 4% 増加して 45,220 件となっています。これらの数値からも、このグ

2. 知財業界における翻訳の問題

知財、特に特許・実案はその国の産業の発達に直接関係するため、各国政府には自国のニーズを勘案しながら、国際法に反しない限度で、独自に知財に関する権利を認可・管理したい要求が普遍的に存在します。例えば、エネルギー、製薬、電気通信機器製造などに関して自国企業の開発、発展が遅れている場合、他国企業が国内市場に参入して独占支配することは避けたい国は数多く存在します。しかし、他国企業による国内市場独占の防止は、強制実施権制度や独占禁止法でも対応可能です。そして、各国特許庁には、先行技術調査、審査、登録管理に関する負担を軽減する必要性

が常に存在します。知財を認可・管理する特許庁を維持するためには、有能で熟練した人材を育成する必要があり、組織を構築・維持するために莫大な負担がかかるからです。

一方、知財制度を利用する企業側には、事業のグローバル化に伴い、世界中で同一の知財を確保し、一元化管理する必要性が高くなってきています。さらに、成長著しい産業では、企業や事業部門の移転・合併・併合・吸収などが活発ですので、知財権利者の名称・住所変更や知財権の譲渡に伴う登録変更手続きの簡素化問題があります。

このような現状から、各国特許庁や知財制度を利用している企業の両方から、上記負担軽減や問題解決を目的として、異なる国での重複手続きを解消する世界共通包袋構想が提唱されるようになりました。後記で詳細に述べますが、近年ディープラーニングなどの人工頭脳 AI 技術の発達により機械翻訳の品質が飛躍的に向上していますので、上記世界共通包袋の構想では、包袋の事務的処理は英語で行うことが前提になっているようです。

ここで、知財の専門家には常識ですが、特許出願書類においては、特許請求の範囲のみならず、権利化後の明細書全体も権利を確定するための重要な役割を担っています。特に、図面で表しにくい技術的思想が発明となることの多い化学発明やコンピュータ発明においては、発明の特定や権利範囲の確定では、明細書の文言の一字一句が重要になります。効果の予測可能性が低いために用途効果の発見が発明の基礎となりやすい化学分野の発明においては、新規性・進歩性のみならず記載要件を検討するために、出願書類全体の文言は重要です。審査において、出願書類に記載されていない文言で補正を行うことは困難ですし、権利化後の訂正ではほとんど不可能です。更に、化学分野の発明は、1の製品に関係する特許の数が比較的小さいため、1の特許の価値が高い傾向があります。例えば、重要な1の特許を有する製薬会社は、その特許でカバーするブロックバスター医薬を世界的に独占販売できます。そのため、出願書類の翻訳に万一誤りがあった場合、権利範囲の制限や権利無効などの致命的な結果が生じる可能性もあります。

一方、機械・電気分野では、1の製品に関係する知財の数は膨大です。巨大企業であれば、大量に保有する特許群に基づいて有利な立場を確保し、クロスライセ

ンス交渉することにより、相手の特許を自社の事業に対して無力化する手段も選択できます。標準技術を利用することが必須の産業では、標準技術策定委員会で認定された標準特許を保有している企業は、強い交渉権を持ちます。なので、たとえ製品全体の差し止めが可能な特許発明であっても、特許の価値があまり高くないこともありえます。また、機械や回路関係の発明は、図面を根拠に補正・訂正が可能な場合が多いので、権利化後の特許請求の範囲や明細書の翻訳に万一誤りがあった場合でも、特許庁への手続きや他社との交渉でリカバリ可能な場合も多いと聞きます。このような、翻訳内の誤記はあまり重要視しなくても良い立場の人々からは、上記世界共通包袋の内容を実質的に英語で運用して、自国特許庁での国内審査は原語で行い、他言語を公用語とする他国特許庁からの出願は、庁同士が共同して機械翻訳を調査、審査すると、特許庁出願人双方にとって効率が良いのではないかと、この提案もあるようです。具体的には、米国、欧州、その他の特許庁において、関係国同士の合意の下に、日本語書類のみで公用語による対訳の無い出願であっても受け入れて（日本特許庁からの協力はあるかもしれませんが）機械翻訳に基づいて審査を行っても良いのではないかと、この提案です。

優位な交渉権を有して世界的事業を行う大企業にとっては、保有知財全体から考えると1の知財の価値は小さいことが多いし、紛争が起こるとしても、相手と対等以上の立場で交渉解決の道が図られる機会が与えられています。さらに、大企業は、多種多様な情報・経験・投下資本を有する知財強者であり、機械翻訳の利点も欠点も熟知しているので、重要な発明を出願する際は正確な翻訳を用意するのが一般的なので誤訳による問題は発生しにくいといえます。一方、中小企業、学術的研究機関、個人などの知財弱者にとって、知財に関する情報・経験・資本は限られています。そして、数少ない知財権しか保有しない中小企業らにとって、1の知財の価値は非常に大きいものですが、往々にして、交渉相手との関係では、情報・経験・資本の少なさから不利な立場となることが多くなります。費用削減が必須の知財情報弱者が、出願国相手先の特許庁における審査対象として人間によるチェック無しの機械翻訳を選んだ場合、権利取得不能、または必要とする権利を取得できない虞れがあります。

このように、人間によるチェック無しの機械翻訳で

も実体審査の対象とみなすことのできるシステムは、各国特許庁と、一部の出願人、権利者にはメリットがあるかもしれませんが、誤訳が致命的になりうる分野の企業や知財情報弱者にとっては不公平な結果を生じる可能性があるといえます。

3. 特許関係書類の翻訳

(i) 翻訳に求められる要件

特許(実用新案を含む)出願書類に関して、文章で権利を特定する「特許請求の範囲」はもとより、「発明の詳細な説明」部分も権利範囲を確定するために使用され、権利書としての役割を担います。そのため、日本以外の国へ出願する場合や、外国から日本へ出願する場合には、出願書類の正確でわかり易い翻訳が求められます。審査段階において発見された誤訳を解消するためには、出願書類の訳文に根拠が見つかるなら補正が行われ、そうでなければ補正より手間も費用もかかる誤訳訂正が行われます。一方、権利化後に致命的な誤訳が発見されて、訂正要件を満たす場合は、更に手間と費用と時間のかかる訂正審判を請求する必要があります。しかし、登録された権利を信頼している第三者が不利益を被らないために、訂正審判が認められる要件は非常に厳しくなっています(特許法第126条:①特許請求の範囲の減縮, ②誤記または誤訳の訂正, ③明瞭でない記載の釈明, ④他の請求項の記載を引用する請求項の記載を当該他の請求項の記載を引用しないものとする事, のいずれかであって、実質上特許請求の範囲を拡張し、または変更するものであってはならない)。必要な誤訳訂正ができない場合は、権利範囲が制限されたり、無効となってしまう虞れが生じます。

上記のように、特許出願書類の翻訳では、訳文が、主語と述語が対応し、自動詞、他動詞の区別ができていくちんとした文章とすることはもちろんのこと(語学)、権利書としての正確性・明確性を備え、将来の紛争の原因とならず疑義の生じない適切な翻訳であって、万一他国のファミリー特許で裁判に引用されても出願人の不利にならない訳文が求められます(内外法律面)。さらに、最先端の専門的技術である特許実案は、その分野のみで使用される特殊な用語、言い回しが多いため、専門用語を適切に使用して、発明の課題とその構成技術を理解できるように翻訳する必要があります(技術面)。また、クライアントが満足する

費用で期間内に納めることも要求されます(実用面)。

(ii) NIPTA (NPO 日本知的財産翻訳協会)

特許出願書類の翻訳を外注する場合は、知財法務に関する知識は別として、少なくとも語学力と技術理解力のある翻訳者を選ぶ必要があります。従来から付き合いのある信頼のおける翻訳者がいない場合、翻訳者を選ぶ指標として、特定非営利活動法人(NPO)日本知的財産翻訳協会(NIPTA)の行う検定試験制度を利用することもできます。⁽²⁾

NIPTAは、特許出願書類などの知的財産に関する翻訳能力を客観的に評価し認定する検定試験を年2回行っています。この検定制度は、発注者側にとっては、翻訳者の知財翻訳能力を客観的に示す公正な指標となります。後援は、特許庁、日本弁理士会、一般社団法人発明推進協会、一般社団法人日本国際知的財産保護協会(AIPPI)、一般財団法人知的財産研究教育財団、一般財団法人経済産業調査会、その他二社です。

NIPTA 検定試験は、インターネット設備のあるコンピュータがあれば自宅でも受験することができ、実際の翻訳作業の環境と同様に、試験中のネット検索やコンピュータ辞書などの利用も自由です。英語の検定試験は、年2回(春季和文英訳)(秋季英文和訳)で、1級は、(1)知財法務実務、(2)電気・電子工学、(3)機械工学、(4)化学、(5)バイオテクノロジーの5つの技術分野を選択して行われます。受験者は、試験時間内に解答のテキスト文章をEメール送付します。1級合格のためには、最終面接を通過する必要があるため身代わり受験することはできません。2級および3級では技術分野の選択はありません。3級は、一部選択式問題があります。中国語またはドイツ語の検定試験は、秋季に中文和訳、独文和訳が行われます。技術選択はありません。

全くの私見ですが、NIPTAの1級試験は、知財法務実務以外は特許出願書類全体を試験対象にしていますし、知財法務実務は契約関係や特許審査関係の書類を含むことがあり、難易度はかなり高いと思います。なので、語学力と技術理解力を兼ね備え、クレームなどの特許特有の言い回しに対応できる翻訳者が、1級受験者の多くを占めているようです。毎回の合格者数は、各分野で0~数人程度です。合格人数に限定は無く、基準を満たしていれば何人でも合格できますので、特許翻訳に携わる者として、一人でも多くの受験者が1級合格されることを願う次第です。NIPTAの

1級合格認定者リスト（和文英訳、英文和訳）では、認定者本人が承諾した範囲内で情報が公開されています。⁽³⁾

なお、外注先から納品された翻訳には、日本、米国、欧州などの実務に適合しているか、権利を限定して解釈される虞はないかなどの観点から、内外法の専門家である弁理士の確認が必要でしょう。

(iii) 機械翻訳の活用法

英語力に自信があり、技術に詳しく、翻訳にかかる時間とコストを削減したい弁理士にとっては、Google（登録商標）翻訳などの機械翻訳を活用することも便利な手段です。機械翻訳はAI技術の進化に伴い発展してきました。AIの発展については後記することにして、ここでは、ネットから容易に無料で利用できるGoogle翻訳の活用法について説明します。なお、Google翻訳は、公開されたネットを利用するものですので、秘密情報に関して利用する場合には注意が必要です。

Google翻訳（<https://translate.google.co.jp/>）はGoogleが提供するサービスで、サイトの例示画面を下図に示します。画面の左欄に入力したテキスト文字を、右欄のプルダウンで選択した目的言語へ即座に翻訳して表示する機能を有します。入力出力ともコピーペースト機能を用いることができるので、簡便に利用できます。原則ほとんどの言語同士で翻訳可能ですが、一般に、英語を介して利用すると精度が上がります。



Google翻訳は、独自に開発された統計的機械翻訳エンジンを利用しています。その翻訳エンジンでは、二千億程度の語から成る国連の正式な文書と国連の翻訳者の訳をコーパス（自然言語処理の研究に用いるため、自然言語の文章を構造化し大規模に集積したもの）として利用し、原文と訳文とからパターンを探り、そのパターンを利用してエキスパートシステムが作成

されています。⁽⁴⁾ 翻訳のための資料として、ネットや各種協力先から入手できるビッグデータが使用されています。Google翻訳は、2016年11月から翻訳アルゴリズムがニューラルネットワークを使用したものに変更され、翻訳精度が飛躍的に向上して読みやすい文章になりました。⁽⁵⁾ 従来のアルゴリズムとは異なり、1の文章をパーツごとではなく、入力された全体として扱う傾向が強くなり、より正確な訳語が得られるようになりました。特に英語を含むインド・ヨーロッパ語圏の言語は共通性が高いため、精度が高いようです。

2010年11月から、欧州特許庁EPOは、複数の言語の情報に対するアクセシビリティを高める目的でGoogleと提携し、Googleは翻訳された特許情報を利用して自身の機械翻訳技術を改良しています。⁽⁶⁾ 機械翻訳は、質の良い対訳ビッグデータにより品質が向上しますし、特許分野では長期間大量の対訳ビッグデータが利用可能な形で蓄積されています。Google翻訳は、語彙力に強いので、大量に列記された化学物質名でも、比較的正確に素早く翻訳してくれます。なので、将来実現するかどうかは不明ですが、知財分野に特化したGoogle翻訳が提供された場合、より精度の高い特許翻訳が容易に入手できると思われれます。

しかし、Google翻訳を利用するには、Google翻訳の問題点を認識する必要があります。まず初めに、Google翻訳へ入力されたデータは、Googleで保存管理され、万一の場合は流出する可能性があります。なので、秘密情報が簡単に特定されるような態様で入力することは避けるべきです。

Google翻訳は、サーバ、ネット環境、コンピュータの処理能力に依存して、1回の処理量に限界があります。長い文章や段落を1回に処理できない場合は、分割する必要があります。ところが、1回の処理内容でも複数の単語が含まれる場合や、異なったタイミングで同じ単語を含む文章を処理した場合は、同じ単語に対応する訳語が同一でないことがあります。

さらに、2016年11月以降のシステムは逐語型でないため、文章の一部が省略されたり、統合されたり、ニュアンスの異なる訳語となってしまう、正確な翻訳が得られない場合があります。上図の具体例では「独自に開発された」がproprietary（独占的に保持した状態の）一語で訳されており、より正確な逐語訳であるindependently developedとはなっていません。一方、「統計的機械翻訳エンジン」は、正確な用語で訳さ

れています。また、例えば、冗長な文章「ボタンを、上、上、下、下、左、右、左、右、B、A、上、上、下、下、左、右、左、右、B、A、上、上、下、下、左、右、左、右、B、Aの順に押す。」を入力すると、同じ訳語が追加されたり省略されたりします。^{(7)143p.}

Google 翻訳の構造上の問題ですが、収集されたビッグデータに依存して、様々な言葉（固有名詞、専門用語等）が全く関係ない言葉に翻訳される可能性が常に存在します。ニッチで蓄積データが少なかったりする分野や、新規に開発され発展中の技術分野であって対応訳語が確立されていない分野では不適切な訳が多くなります。

Google 翻訳サイトは性善説に従い運営されていますので、悪意の利用者が「情報の修正を提案」機能を利用して誤訳を学習させることも可能です。不正確な対訳データがウェブ上に大量に開示されて翻訳精度が低下する虞れもありますし、政治的に蓄積データを不正に利用されたり変更されたりする可能性すらあります。しかし、問題が起こっても、Google 翻訳はビッグデータに基づく複雑な AI 技術のブラックボックスですので、詳細な検証はできません。

以上の理由から、Google 翻訳は、正確性が必須の要件である特許出願書類の翻訳において、あくまで翻訳の補助手段として利用するべきです。しかし、翻訳の補助手段としては非常に有効なツールです。下記(a)～(e)に、機械翻訳を利用して精度の高い訳文をより簡便に作成する方法を説明します。

(a) 原文を、機械翻訳しやすい日本語で作成します。特定の技術分野でのみ使用される固有の慣用的表現は避け、平易で直接的な表現で一般人でもわかり易い文章にしましょう。だらだらと続く長文は避け、「しかし」、「一方」、「以上の結果」などの接続詞を使って、適切な文にしましょう。文章の主語に対応する自動詞または他動詞か、なにに対する目的語か、係り受けは正確か、修飾関係は明確か、などの確認も大切です。代名詞（その、この等）の示すものが直接的にわかり難い場合は、具体的な言葉を補いましょう。日本語特有の言い回し（「かまぼこ形」など）は避け、万国共通の表現にしましょう。文章の流れで理解できても、その一文だけ抜き出すと別の意味を表す場合もあります。「十分反応させる」からは、「充分反応させる」または「10分間反応させる」の二種類

の意味が生じます。内容が一義的に定まるように、例えば、様々な意味を表す「・・・で」は避けて、「において」、「を使用して」、「の間」など具体的な表現にしましょう。⁽⁶⁾

- (b) 技術者である発明者本人が明細書を作成するなど、作成段階で上記注意が困難な場合でも、原文をチェックして、接続詞を挿入して長文を区切る、主語を補う、代名詞を具体的な言葉に代替するなどのプレ翻訳操作を行うと、精度の高い機械翻訳が得られます。
- (c) 複数回使用する用語、語句の訳は、統一されなければなりません。機械翻訳前に一括置換する、機械翻訳後に対訳リストに基づき確認するなどの対処が必要です。下記(d)秘密性を考えると、機械翻訳前に、キーワードの対訳にそれぞれ暗号を付与し、対訳リストを利用してそれぞれの暗号へ一括置換しておき、機械翻訳後に、対応する訳語に置換するとよいでしょう。
- (d) 従来技術ではなく、具体的な秘密情報に関する内容を入力する場合は、固有名詞やキーワードを一括変換し、翻訳対象を分割し、順番を変えたり時間をおいて入力すると、秘密担保性が高まります。
- (e) 機械翻訳後は、訳文中に抜けはないか（特に数字、近傍にある同じ表現など）、言語的にも技術的にも内容は正しいかを逐語的に確認します。英語の基本である、冠詞の区別、単数複数、自動詞他動詞の違い、動詞の時制や受け身に気を付けましょう。技術用語は、信頼できる辞書や、信頼できるネットサイトなどで確認します。たとえ広く使用されているネット辞書やサイトであっても機械翻訳であったり、英語を母国語としない著者のものであったりして不適切なこともあります。参考資料の作成者の名前や国籍を確認することも有益です。技術分野が違えば、同じ言葉でも異なる意味を表す場合があります。法律、審査基準、判例など法律的な観点からも確認しましょう。

上記(a)～(e)は、出願書類の和文英訳を念頭に置いていますが、特許庁関係の書類や通信文でも利用できます。ただ、庁関係書類には長年使用されてきた特有の言い回しが多く、直訳では理解困難な場合もありますので、相手にとってわかり易い翻訳にする必要があります。一般的な定型訳文も利用できますが、文言や

条文番号の表現が一致するように気を付けましょう。特殊な法律用語、技術用語は、信頼できる辞書などで確認します。

4. 意匠・商標関係書類の翻訳

「意匠にかかる物品」や、商標の「指定商品」「指定役務」の項目も権利範囲に直接的に影響するので、その翻訳には高い精度が求められます。対応する訳語を単純に当てはめるのではなく、翻訳対象を理解して、内容を正確に訳しましょう。上記 Google 翻訳も便利に利用できます。しかし、新規で斬新な事業分野の商品・役務を指定する場合、特に区分や訳語に気を付ける必要があります。将来的にもクライアントに有利となる権利範囲をできるだけ広く確保すると共に、オフィスアクションを受けず、不使用とならないよう、別表の文言、登録例などを参考に前例のある明確な訳語を選ぶべきです。1 区分内の同種の指定商品があまりに詳細かつ多数記載されていて、出願相手国の規則を遵守しながらも大概念の指定商品を記載可能な場合は、クライアントの了解の下で、簡略化記載することにより翻訳の手間と費用を削減することもできます。

5. 英語での発表

渉外実務を担当する弁理士であれば、英語で発表する機会もあると思います。最も重要なのは、当然ですが発表内容とその構成です。聴衆が期待する内容を、わかり易いストーリーの日本語で作成してから翻訳します。スペルチェックは勿論ですが、英語の基本（冠詞、単数複数、自動詞と他動詞、受け身、熟語に含まれる前置詞の区別など）に気を付けます。可能なら、ネイティブチェックを受けることも良いですが、技術系より法律系のネイティブの方が言語に気を使ってくれるようです。スライドは、印刷しても会場発表でも見やすい文字で作成します。単調にならないように、図表以外にも、ロゴやイラスト、ピクトグラムなどを挿入しても良いでしょう。資料に使用する色の組み合わせを統一すると、好感度が増します。最初に、主催者や直前発表者に言及したり感謝の意を表したり、聴衆の興味を引くトピックを述べてから始めると、場が和みます。できるだけ前を向いて、聴衆の目を見ながら自信を持って発表しましょう。

6. 人工知能 AI について

機械翻訳は AI 技術に基づくものですので、その限界と問題点を知るために、AI についても簡単に説明します。

AI 技術は、第一次 AI ブーム（1950 年代後半～1960 年代）、第二次 AI ブーム（1980 年代）、第三次 AI ブーム（2000 年代以降）と段階的に発展してきました。

第一次 AI ブームは、限定された条件下での推論・探索技術を磨いて問題を解く、チェスなどのゲームへ適用できる技術の開発です。推論・探索するために、深さ優先や階層の幅優先で探索する技術が発達しました。第一次 AI ブームの目標を達成した例として有名なのは、1997 年にチェスの世界選手権優勝者ゲイリー・カスパロフを破った IBM のディープブルー、2011 年に米国クイズ番組で人間チャンピオンを破った IBM ワトソン、そして 2012 年に将棋電王戦で米長邦雄永世棋聖を破ったボンクラーズです。これらは第二次ブーム以降の、より知能的な問題処理能力、高速計算能力、莫大な量のデータマイニングの結果、達成されました。AI と人間との勝負では、選択肢が狭まってくる中盤以降に AI が強さを発揮し、最終局近くでは AI はノーミスになります。このような高性能のゲーム用 AI は、機械学習の適用、確率的な有利性を示す指標である特徴量（ビッグデータから抽出される特徴を定量化したもの）の発見、スコア評価を高性能の演算処理能力を利用したモンテカルロ方式にすることなどにより飛躍的に発達しました。しかし、ゲームに特化したディープブルーやボンクラーズは確率計算を出力しているだけですし、ワトソン IBM の質問応答システムも、質問の意味を理解しているわけではなく、質問中に含まれるキーワードに関連する確率の高い答えを出力しているだけです。なので、推論・探索技術へ機械学習を組み合わせても、現実に求められている複雑なフレーム問題の解決は困難でした。

第二次 AI ブームは、専門的知識を学習させて問題を解決するエキスパートシステムの発展により起こりました。限定された特定分野の情報を入力した AI から、自動運転により解答を得るもので、現実の産業分野でも広く応用されています。天気予報など膨大なデータに基づく予測や、プラント設備内などでの測定値から異常を検知するシステムが開発されましたが、知識の発見や意思決定のレベルまでは達していません。エキスパートシステムを発展させて、ジャパン

ネット銀行では、2016年にAIによって完全自動化したローン与信審査を行っています。

しかし、専門家から技術情報などの知識をヒアリングするのは大変な手間、時間、費用がかかります。そして、あいまい表現や数値化できない表現を含む知識情報を、入力記号の形で整理し、情報の関係をルール化して適切に維持管理してAIへ学習させ、推論を行わせるのは困難な作業です。人間の持つ一般常識を入力することを目標として1984年に発足したCycプロジェクトは、30年かかっても作業が完了していないようです。

オントロジーとは知識記述の研究であり、知識を上位下位関係や包含関係により概念化して明示的な仕様とする研究です。第二次AIブームのエキスパートシステムは、人間が考えて知識を入力するヘビーウェイトオントロジーでした。それに対し、1990年代から統計的自然言語処理技術が進化し、コンピュータへデータ入力して自動的に概念間の関係性を発見するライトウェイトオントロジーが発達しました。ライトウェイトオントロジーは、完全な正確性を求めず機械的処理するデータマイニングと相性が良いものです。

また、1990年代半ば以降は、インターネットが爆発的に普及し、2000年代には機械学習が広がりました。機械学習は、AI自身が、データベースなどから、サンプルデータ集合を入力して解析を行い、そのデータから有用な規則、知識表現、判断基準などを抽出し、アルゴリズムを発展させるものです。データ集合の解析には統計処理が用いられます。機械学習には、訓練学習（入力と正しい出力がセットになった訓練データを用意して、AIに学習させる）と訓練無し学習（AIへ入力データのみを与え、データ内のパターンやルールを機械的に抽出させる）があります。主なデータ判断法には、最近傍法、ナイーブベイズ法、決定木、サポートベクターマシン、およびニューラルネットワークがあります。ニューラルネットワークは、人間の脳神経回路を模倣しており、単純3層構成の場合、入力層、重みづけ、隠れ層、重みづけ、出力層で処理し、誤答したら選択肢の重みづけを変化させて認識の精度を上げています。ニューラルネットワークの学習段階は時間がかかりますが、予測段階の処理速度はとても速いものです。

機械学習では、目的とする対象に関する特徴を入力データ内から選択して、高い精度を有するように特

徴量を決定する特徴量設計 (Feature Engineering) を行います。しかし、この特徴の選択と特徴量の決定は、実際は、人間の最終的な微調整と確認が必要なので手間がかかり、ニューラルネットワークで高階層にすると、重みづけの変化が最終出力に影響しにくくなるため精度上昇がみられず、単純モデルにしか使えない欠点がありました。

第三次AIブームは、ディープラーニング（特徴表現学習）の開発により起こりました。ディープラーニングは、ビッグデータを利用する機械学習において、データ判断にニューラルネットワークを使用する画期的な技術であり、2012年にトロント大学チームが発表しました。ディープラーニングでは、データ内の特徴表現を認識して分析するアルゴリズムを、AIが自ら発展させます。高階層で構成され、1階層ごとに学習して答え合わせをするオートエンコーダで情報を圧縮し、各層で重みづけを修正して、特徴表現が抽出されます。ディープラーニング技術では、特徴量や概念の頑健性（ロバスト性）を高める技術も開発され、最近ではほとんどのAI技術に利用されています。初期の具体例として、2012年にGoogleは、ネットから入手した1000万枚の画像処理を、実質1万6000個のプロセッサで3日間処理して猫の概念を抽出し、認識しました。なお、その画像から抽出された概念が、言語（記号）としての「cat」で表現されるために関係づけを行う記号設置は、人間が教える必要があります。非常に困難ですが、実世界の概念を言語（記号）と結び付ける記号設置問題を解決して、ディープラーニングと組み合わせれば、データ内の概念を認識して言語に置き換えることができ、高い精度の機械翻訳が達成できると期待されます。

しかし、ディープラーニングには、現在の環境では解決できない問題があります。例えば、2011年に発足した「ロボットは東大に入れるか」(東ロボくん) プロジェクトでは、東ロボくん AIに東大入試レベルの英語問題を解答させる試みを行いました。⁷⁾ 東ロボくんは、2016年に500億単語からなる19億文をディープラーニングで学習しました。その結果、提示された単語を問題で設定された場面に応じて適切な文章にする語句整序問題は、ほぼ100%の精度で解答しました。しかし、更に大きなデータを入手して150億文を学習しても、東ロボくんは、適切な会話を4つの選択肢から選んで会話のやりとりの間に挿入する会話完成問

題の正答率は、4割以下でした。論旨要約などでは、ディープラーニングの東ロボくんは、既存の手法より悪い結果でした。つまり、東ロボくんにとって、問題文に明記されていない、会話のやり取りの前提とされているような一般常識に基づく問題は難しかったのです。一般的な様々な常識を、整合性がとれるように組織的にルール付けることは、Cycプロジェクトで試行されたように非常に困難ですし、機械学習させるためには現在入手可能なデータでは足りません。東ロボくんのディープラーニングに必要なデータは、広く標準的に使われている「お手本になるような日本語と英語の対訳」です。しかし、普遍的で正確な文章データには限りがあり、今回使用したデータ量150億文を飛躍的に大きくする手段は見つかりません。ネット上に存在するデータは、特殊性、不正確性、刹那性、不適切性などの欠点があるため、安易に利用できないからです。

Siri（登録商標）やGoogleアシスタントも、基本的には、検索して得られた言語情報を統計処理して、正答率の高い回答を導き出すディープラーニング技術を利用していますので、検索対象となるデータに依存します。大量の適切なデータを用意できたとしても、意味を理解していないので正答率100%は達成できません。しかし、対話型AIの利用者は、AIの回答を自分の求める内容に近づけるように好意的に解釈する傾向があり、あいまいな回答であれば対話を続けようとし、AIが議論をそらしたり、言いよどむ回答をすると人間らしく感じて満足してしまいます。なので、AIの対話技術では、高度な正確性は求められないようです。

東ロボくんプロジェクトで、統計的手法ではなく自然言語処理技術を利用するには、文法などの言葉のルールを覚えさせ、論理的、演繹的手法で翻訳精度を向上させる必要があります。まず、問題文の構造を正しく解釈するために、文節に区切り、主語と述語を特定し、述語と目的語の関係や修飾と被修飾の関係を決定します。しかし、正確に特定・決定するためには、対象とする文章中には記載されていない一般常識が必要になる場面が多く、常識を学習していないと誤訳が発生します。例えば、区切られた分節中の主語候補として人間を表す単語と無生物を表す単語がある場合であって、その述語の主語は人間しかありえない場合は、候補の選択を誤らないように明確なルールを作成

し、擬人化の場合分けも含めて学習させる必要があります。このように、論理的な自然言語処理を行うためには、高度で精緻に構築された膨大な言語ルール体系の設計が必要となりますし、その適用には高い処理能力を備えたコンピュータも必要となります。

以上より、ディープラーニング技術は、多層ニューラルネットワークに関する技術であり、基本的には統計処理技術を用いた機械学習です。なので、質の良いビッグデータを用意できる限りにおいて有用であり、精度の高い結果を得ることができますが、統計処理技術を用いる限り、100%正確な結果は望みえません。正確な結果を期待するには不向きなデータとは、量が不十分、多種の例外がまれに存在する、ルールが明確でない、言外の意味が生じる可能性がある、変動が激しく調整が間に合わないデータですが、機械翻訳に現在提供可能なデータにはかなり当てはまってしまう。将来、一般常識まで認識できるレベルの適切なビッグデータが用意でき、記号設置などの関連する問題も解決するブレークスルーがあれば、機械翻訳の精度はかなり上昇するでしょう。^(8,9,10,11)

特許文献の機械翻訳に関して、過去約50年間（1965～2016年）の累積で、日本から米国への出願が約200万件、米国から日本への出願が約100万件、合計約300万件の対訳データが存在します。なので「特許文献については人手による対訳データが大量に存在しているので、特許文献の自動翻訳が実用レベルに達するのは時間の問題である。」との意見もあるようです。⁽¹²⁾しかし、出願される技術分野は非常に広範囲であり、その分野によって出願数の偏りがありますので、かなりの分野ではディープラーニング技術に必要とされるデータ量は不十分であると思われます。さらに、現在のディープラーニング技術では論理的な自然言語処理は行われていませんので、意味は理解せずに統計処理技術を用いて翻訳する場合、100%正確な訳文は担保できません。

そもそも、特許文献でクレームされている発明や考案は、本来今までになかった新規で進歩性を有するものですから、たとえ公知要素の組み合わせであっても、ビッグデータ内に記載も示唆もされていない、ビッグデータからは容易に想起できない思想のはずです。たとえば、常識からは考えつかない組み合わせを構成要件とする発明、公知物質の従来知られているとは別種の効果を利用する発明などを翻訳する場合、

従来の技術を集めたビッグデータに依存するディープラーニング技術を採用する限り、クレーム発明を正確に翻訳する確率が低くなるのは当然とも考えられます。

AI とビジネスとの関係について付け加えます。AI の発達により将来無くなる可能性の高い職業の一つとして弁理士がある、と一般に広まっているようで、その根拠は、2015 年に発行されたイギリス・オックスフォード大学と野村総研の共同研究の報告書にあるようです。⁽¹³⁾しかし、上記説明の通り、ディープラーニング技術はビッグデータに基づく統計処理であり、意味を理解する AI では決してありません。なので、顧客から相談を受け、発明を抽出し、従来技術とその課題を調査してクレームを作成し、出願書類の形にまとめ出願する手続きを、現在の延長線上の AI が代替することは不可能です。勿論、特許庁や裁判所における顧客の代理やコンサルタント業務もできません。さらに、AI にはビッグデータが必要ですが、機械翻訳のための対訳データでさえ限られているのに、判例数ともなると、日本はもとより米国でさえ充分ではありません。上記報告書中の「AI に代替される可能性が 92.1%」の根拠は、研究に参加したオックスフォード大学のマイケル・オズボーン准教授の関連論文が根拠らしいと言われています。知財に関する米国裁判では、Disclosure (文書開示) 手続きにおいて提出された e-メールなどの膨大な電子書類内から裁判に関係する書類を発見する作業は、特許弁護士ではなく、主にパテント・エージェントが担当しているらしく、その作業は膨大でかなりの費用がかかっています。このような電子のデジタルデータを相手にする作業は、個々の案件で異なるキーワードやパターンを抽出し、汎用化されたディープラーニング技術を調整して機械学習させた AI により代替可能となるであろうことは納得できます。パテント・エージェントは、米国特許商標庁が行う特許のみの択一選択試験に合格して、米国特許商標庁に対して特許出願手続き代理業務を行うことができる米国内の資格者です。日本弁理士のような条文理解は必要ありませんし、代理権は対庁特許出願手続きに限られます。察するに、オズボーン准教授の関連論文は、出願手続きをあまり扱わず Disclosure 手続きで証拠資料を発見する作業を主な業務とする一部のパテント・エージェントを対象とするものであり、日本側共同研究者による弁理士業務の理解不足も加わり、

日本の弁理士の仕事は、Disclosure 手続きを主な業務とするパテント・エージェントの仕事と同じであるとの前提で分析し、報告を作成したものと思われます。

AI 技術が導入できるのは、弁理士業務の中でも事務作業です。しかし、特許商標事務所の事務内容に関してディープラーニングに利用できるようなビッグデータは存在しません。ディープラーニング技術を使ったエキスパートシステムを採用する場合、例外が多くて改正が頻繁に起こる特許商標事務を体系化して、汎用 AI へ学習させるためには、案件を多く扱う巨大事務所であっても巨額な費用を負担しなくてはなりません。なので、従前の期限管理などのファイル管理システムを、実際の事務に合わせて適宜改良したり機能を追加する利用方向は、今のところ変わらないと思われます。つまり、一般的な弁理士事務所では、業務が専門的すぎてディープラーニングで AI 化する費用に見合う経済的効果が得られにくいでしょう。一方、大企業の知財部の場合、AI を導入する場所はどこでもよいので、地方や海外で AI により知財事務や会計処理を行い、社内コストを下げることも充分考えられます。世界共通の手続きであれば、ますます国内で処理する必要はありません。AI 操作は英語と親和性が高いので、AI 導入により海外との競争が激しくなると予測されます。

将来的な AI の進歩に伴い、弁理士は、単調な仕事は AI に任せ、より高次の仕事をするのが求められます。初めて発生した問題や、複雑で総合的な問題が発生した場合、経験・知識のある人間しか対応できません。AI に対抗できる業務としては、対人的スキルを必要とする発明の発掘やコンサルタントなどの、大量で精緻なデータなしでも判断できる仕事をする、複数の技術が関係する分野を扱う、顧客の事業戦略に沿って関連事業の調査を行う、弁理士業務に関連する AI システムを構築または運営する専門家になる、などが考えられます。⁽¹⁴⁾

7. まとめ

知財関係者の中には、グローバル化、英語の世界共通言語化、機械翻訳の精度上昇を鑑みて、世界共通袋の内容を実質的に英語で運用して、他言語を公用語とする他国特許庁からの出願は、出願された国の特許庁が機械翻訳で調査、審査できるようにしようとする提案があります。しかし、特許出願書類が適切に審査

されて権利を付与されるためには、翻訳に高度の正確性が求められますので、現在も将来的にも誤訳を含む可能性の高い機械翻訳に頼ることを可能とするシステムは、誤訳が致命的な分野の出願人や、知財情報を知らず、資本も不足する弱者にとって非常な不利益を生じます。機械翻訳は翻訳の補助として有利なツールですので、その欠点を知りつつ、上記(a)～(e)の注意に従って機械翻訳を利用することを提案します。現在のAIブームの基礎であるディープラーニング技術は、目的に適したビッグデータを確保しない限り有効に活用できません。なので、弁理士の関係するビジネスでは、AIの課題を知ったうえで将来を展望すべきでしょう。

筆者が試験委員をさせていただいている NIPTA の事務局、その他沢山の先生から本稿作成へご協力を頂きまして、誠に有難うございました。なお、本稿の内容は、一般に公開されている情報のみに基づいて筆者の私見を述べたものであり、いずれの個人・組織・団体とも関係ありません。

(参考文献)

- (1) 特許庁, 行政年次報告書, 2017 年度版, 第 1 部, pp5~6
- (2) NPO 日本知的財産翻訳協会, “知的財産翻訳とは”, <http://www.nipta.org/>, 日本知的財産翻訳協会ホームページ, (参照 2018-05-18)
- (3) NPO 日本知的財産翻訳協会, “知的財産翻訳検定 1 級合格

認定者リスト”, http://www.nipta.org/NIPTA_Grade1_List/J_Grade1_List_je.html, 日本知的財産翻訳協会ホームページ, (参照 2018-05-18)

- (4) Philipp Lenssen, “Google Translator: The Universal Language”, Google Blogscoped ホームページ, 2005, <http://blogscoped.com/archive/2005-05-22-n83.html>, (参照 2018-06-18)
- (5) Google 翻訳チーム, “Google 翻訳が進化しました。”, Google Japan ホームページ, 2016, <https://japan.googleblog.com/2016/11/google.html>, (参照 2018-06-18)
- (6) 浅見節子, “特許における明晰な日本語の重要性和特許庁の機械翻訳に対する取組”, 産業日本語研究会ホームページ, 2014, <https://www.tech-jpn.jp/wp-content/uploads/2014/08/23.pdf>, (参照 2018-05-18)
- (7) 新井紀子, AI vs. 教科書が読めない子供たち, 東洋経済新報社, 2018, 287p.
- (8) 松尾豊, 人工知能は人間を超えるか: ディープラーニングの先にあるもの, 株式会社 KADOKAWA, 2015, 263p
- (9) 大関真之, “機械学習入門: ボルツマン機械学習から深層学習まで”, 株式会社オーム社, 2016, 201p.
- (10) 樋口晋也, 城塚音也, 決定版 AI: 人工知能, 東洋経済出版社, 2017, 255p.
- (11) 鳥海不二夫, 強い AI・弱い AI: 研究者に聞く人工知能の実像, 丸善出版株式会社, 2017, 257p.
- (12) 高倉成男, 知財管理, vol68, No4, pp430~442
- (13) 真島加代, “「弁理士は AI に代替される確率 92%」に弁理士会が反論! 高度な対人スキルの奥深い職業”, Business Journal ホームページ, 2018, http://biz-journal.jp/2018/02/post_22313.html, (参照 2018-05-18)
- (14) トーマス H ダベンポート, ジュリアカービー, “AI 時代の勝者と敗者”, 日経 BP 社, 2016, 394p.

(原稿受領 2018. 5. 24)