

# 中国・韓国特許調査における 機械翻訳の利用に関する研究

——特許機械翻訳ツールの現状と  
調査業務に即した翻訳ツール評価——

情報検索委員会  
第1小委員会\*

**抄 録** 近年、機械翻訳技術について、ニューラルネットワークによる新方式が開発され、特許文献を翻訳対象とするシステムが利用できるようになった。新方式による翻訳は、従来の逐次翻訳方式による翻訳と比べ自然で流暢な翻訳文を生成するが、原文内の単語の抜けや揺れなどの課題があることも報告されている。

企業の活動範囲がグローバルになるに従い、知財部門の調査担当者はあらゆる外国語の特許文献を迅速に精度よく調査する必要に迫られている。不慣れな外国語での特許調査に対応するため、機械翻訳を業務に利用し始めている企業は多いと考えられる。

本稿では、機械翻訳機能を利用して外国特許調査業務の精度を維持しつつ効率化を図るという観点から、機械翻訳技術の現状を概観し、特に中国・韓国特許を日本語で調査する環境と業務に即した翻訳ツールの評価方法について検証、考察する。

## 目 次

1. はじめに
2. 外国特許調査と機械翻訳
  - 2.1 外国特許調査モデル
  - 2.2 機械翻訳技術の動向
3. 外国特許のサーチ環境
  - 3.1 中国・韓国特許調査ツール
  - 3.2 サーチ機能の検証
  - 3.3 小 括
4. 特許業務に即した評価方法の提言
  - 4.1 従来の評価方法に対する意見
  - 4.2 知財業務に即した評価手法
  - 4.3 評価条件
  - 4.4 評価結果に対する考察
5. まとめと提言
6. おわりに

## 1. はじめに

機械翻訳技術が進展している。2016年秋にWIPO TranslateやGoogle翻訳といったWebサービスでAIベースのニューラル機械翻訳NMT (Neural Machine Translation) の搭載が相次いで発表された<sup>1), 2)</sup>。このNMTは出力する翻訳文が流暢で読みやすいことで注目を集めたが、原文内の単語が翻訳結果に含まれない(翻訳抜け)、同一文書内の同じ単語が異なる語に翻訳される(翻訳揺れ)、あるいは、原文に無い単語が翻訳文に出現する(過翻訳)など、翻訳文を利用する上での課題があることも報告されている<sup>3), 4)</sup>。

\* 2017年度 The First Subcommittee, Information Search Committee

特許分野では古くから機械翻訳の利用がされており、原文と翻訳文を収録してスクリーニングできる特許データベースがあった。当委員会では、過去に中国・韓国特許の調査での翻訳精度について報告している<sup>5)~7)</sup>。最近では、中国・韓国特許を対象として、原語と翻訳語（日本語、一部は英語）でサーチできる特許データベース（CNIPR日本版(以降CNIPR)<sup>8)</sup>、PATBRIDGE<sup>9)</sup>、中韓文献翻訳・検索システム（以降、中韓翻訳システム）<sup>10)</sup>、PatentSQUARE<sup>11)</sup>）も登場している。

企業の知財部門では、活動する国・地域の拡大にあわせ、当該国・地域の特許文献を迅速に精度よく調査する必要に迫られている。欧米言語から中国・韓国を含むアジア言語にいたるまで対応できる体制を整えることは難しい。事業の展開にあわせて、直ちに多くの国・地域の原語に精通する人材を自前でそろえるよりも、INPADOC<sup>12)</sup> や、DWPI<sup>13)</sup> などの英文抄録、EPO Espacenet<sup>14)</sup> やWIPO PATENTSCOPE<sup>15)</sup> などの外国特許庁のデータベースの翻訳機能を利用することが多いと思われる。

機械翻訳技術の進歩により日本語で高品質な調査・スクリーニングできる環境が整いつつある。各企業の知財部門は、特許調査業務に使えるツールをタイムリーに見極め活用すること

で、調査業務全体の質や運営効率を高めることができる。

本稿では、多くの日本企業にとって重要であり、外国特許調査の関心も高いが言語的な難しさもある中国・韓国特許を対象に、日本語で調査するケースを取り上げる。中国は近年の出願件数の増大やイノベーション段階への変化があり、韓国はすでに電気・IT分野などを中心に世界で技術競争をしているなど、周知である。外国特許調査での機械翻訳技術を概観した後、原文と翻訳文をサーチするツールと、スクリーニングでの翻訳レベルと特許調査業務のマッチングとしての翻訳評価について検証した結果について報告する。

## 2. 外国特許調査と機械翻訳

### 2.1 外国特許調査モデル

サーチャが外国特許調査で機械翻訳（翻訳文データ）を利用する場面は大きく2つある。（I）特許データベース内に蓄積された翻訳利用と、（II）翻訳Webサービスのリアルタイム翻訳利用である（図1）。前者では検索クエリを入力し、後者では直接原文を入力するというサーチャ操作の違いはあるが、どちらも外国特許の原文を機械翻訳した結果を表示出力する。

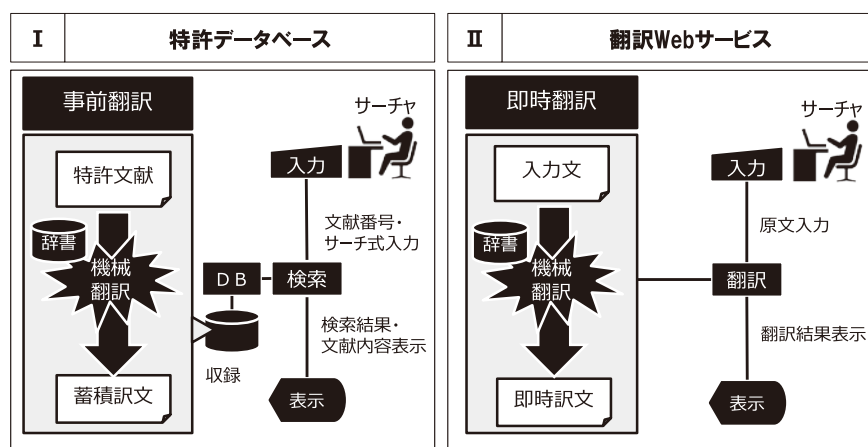


図1 外国特許調査での機械翻訳文利用形態

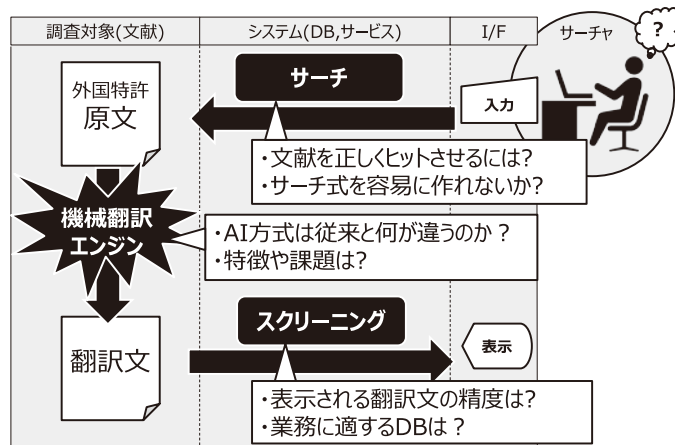


図2 外国特許調査モデルとサーチャの関心

機械翻訳技術を利用した外国特許調査業務を、機械翻訳エンジン、サーチ、スクリーニングの3つの要素を含むモデルでとらえることにした。各要素に対するサーチャの代表的な関心事とともに示す(図2)。

機械翻訳エンジンの動向について次節で述べ、サーチおよびスクリーニングについてはそれぞれ3章、4章で詳説する。

## 2.2 機械翻訳技術の動向

2012年度に当委員会では中日翻訳ツールを評価した。その中で現在もサービス提供しているツールを検証したところ当時に比べ格段に精度が向上していた。AIベースのNMT方式に変更された現在のGoogle翻訳では、人手翻訳に近い翻訳文を出力する(表1)。

表1 Google翻訳文の比較例(携帯電話に関する中国特許公報CN1337812A)

Google翻訳への入力文 中国特許公報 要約部分	本發明可旋显示屏新式移动电话与以往不同是其显示屏相对于主体可以平行旋转。它既容易舒适掌握, 又可以根据观看需要旋转和更换显示屏。解决了今后宽带信息技术要求屏幕尺寸增大与在手机上观看网上影视文艺节目时水平尺寸大于垂直尺寸之间的矛盾。达到了利用现有技术稍加改进, 就能满足今后信息和娱乐发展趋势的需要。
正解とする日本語文 (人手翻訳)	本發明は、ディスプレイが回転可能な新型携帯電話であり、以前と比べて異なるのは、前記ディスプレイが本体に対して平行に回転可能になっていることです。それは、容易且つ快適に持て、そして、ディスプレイの回転と変更に応じて(楽しく)見ることができる。今後の通信(ブロードバンド)技術の要求による携帯電話の画面サイズの増大傾向と、携帯電話でネット上の映画番組を見て楽しむ際の画像の水平サイズが縦サイズより(に対して)大きい、という間の矛盾を解決できた。既存の技術を改良すれば、今後の通信や芸能の発展傾向ニーズを満足できる。
Google翻訳の出力文 (2012年時点での翻訳結果) <sup>7)</sup>	本發明は、回転に平行であってもよい体に対して異なるディスプレイを備えた新しい携帯電話が表示されスピニングしてもよい。それがホールドを取るの簡単で快適ですが、彼らは、ディスプレイを回転させて交換する必要があることができます。IT要件の画面サイズは、水平サイズや電話で文化的なプログラムが増加し、将来のブロードバンドを解決するため、縦方向のサイズとの間の矛盾よりも、オンライン大きく見る。わずかに改善するために、既存の技術の利用の必要性に、将来の情報とエンターテインメントの動向を満たすことができるようになります。
Google翻訳の出力文 (2018年4月時点での翻訳結果)	本發明の回転型ディスプレイの新しい携帯電話機は、表示画面を本体に対して平行に回転させることができる点で従来のものとは異なる。持ち運びや取り扱いが簡単で、必要に応じて回転や交換が可能です。将来のブロードバンド情報技術の画面サイズの増加と、携帯電話でオンライン映画やテレビエンターテインメントを視聴する際の垂直サイズよりも大きな水平サイズとの間の矛盾が解決される。将来の情報やエンターテインメントの発展動向を満たすために、既存のテクノロジーをわずかに改善する必要性を達成する。

機械翻訳エンジンは、開発される世代が進むごとに精度向上が図られてきた。第1世代と呼ばれるRBMT（ルールベース機械翻訳；Rule Based Machine Translation）は文法規則と単語辞書に基づいて翻訳文を生成していたが、第2世代のSMT（統計機械翻訳；Statistical Machine Translation）では対訳コーパス（異なる言語の文と文が対訳の形でまとめられたデータ）による統計処理で翻訳することで、RBMTよりも精度が向上した。そして昨今採用発表が相次ぐ第3世代のNMTは、対訳コーパスをニューラル処理で翻訳することで、SMTを超える精度向上をもたらすと予測されている（図3）。

ここで、機械翻訳の各方式の原理と特徴を表2に示す。

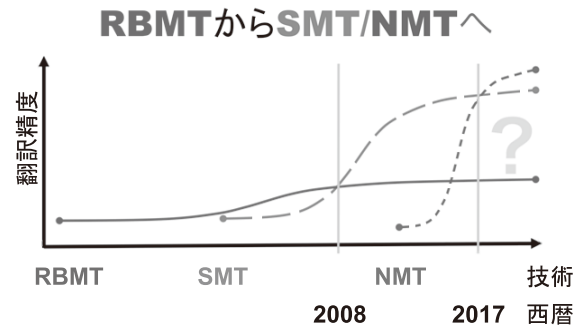


図3 機械翻訳エンジンタイプと翻訳精度<sup>16)</sup>

### 3. 外国特許のサーチ環境

#### 3.1 中国・韓国特許調査ツール

機械翻訳機能を搭載した中国・韓国特許調査ツールについて考察する。

表2 機械翻訳方式の原理と特徴

方式	ルールベース機械翻訳 (RBMT) (第1世代) 1950年代から	統計機械翻訳 (SMT) (第2世代) 1990年代から	ニューラル機械翻訳 (NMT) (第3世代) 2010年代から
原理	人手により作成された規則によって原言語文の構文を解析し、原言語から目的言語へ構造的変換する。	大量の対訳コーパスを用いて、語の対応や語順の並べ替えを統計的に学習し、それに基づいて翻訳する。	翻訳機械に学習機能をつけ、膨大な対訳データをコンピュータ自身が「学習」し、ニューラルネットワークを用いた計算モデルから訳文を生成する。
開発の歴史	当時のマシンスペックでは、言語間の構造のずれを人手の規則に沿って扱うことは難しく、精度の高い翻訳の実現は難しかった。	1991年にIBMが代表的な論文を発表。インターネットを活用することで大量の用例を集めることが可能になり、それを使った「統計ベース」の機械翻訳が登場した。	NMTは、データ処理能力と速度が格段に進歩した現代だから可能になった技術。2016年にGoogleがGoogle翻訳にNMTを導入することを発表したことで、NMTが身近な技術となっていった。
翻訳文の特徴	ルールに合った翻訳がなされると非常に質の良い翻訳ができる。ルールに合致しない場合こちない翻訳になる。	訳抜けや漏れはほとんどなく翻訳の質が安定している。翻訳文としては逐語訳的である。	文章としては非常に自然な訳出をする。翻訳抜けや過翻訳（原文にない語が訳文に現れる）が発生することがある。
課題	辞書と文法の開発に言語の専門家を要し、膨大なコストが辞書・文法開発に費やされる。	日本語と英語のような文法構造が異なる言語間では、翻訳の精度が低く、ポストエディットと呼ばれる後編集が必要。	計算モデルを用いて翻訳を行うため、人間ではやりようのない大きな間違いをすることもある。訳が自然なため、翻訳抜けに気づけない。



近年の機械翻訳技術の進展とともに、自然文とは異なった特殊な言い回しや表現が用いられる特許の世界にも、機械翻訳技術が適用されるようになってきた。中国語・韓国語に精通していなくとも、サーチした結果を機械翻訳された日本語で閲覧することで、手間の省略や時間の短縮といった恩恵を受けた経験をもつサーチャは多いと思われる。

最近になって、翻訳語がサーチ結果の閲覧だけでなく、サーチ時にも活用されるようになってきた。多くのデータベースでは、原語および／または翻訳語を検索語として、データベースのインターフェースから検索することが可能となっている。

今回、各データベースが提供する特徴的な機能について、サーチの精度向上・効率化にメリットがあると考えられる機能の検証を試みた。検証は次の3つの観点からアプローチした。(1) 中国語・韓国語の各原語の特性に配慮した機能からのアプローチ、(2) 機能メンテナンスによる検索精度の向上からのアプローチ、(3) 検索キーの展開や拡張といった支援機能からのアプローチ、である。

### 3. 2 サーチ機能の検証

#### (1) 中国語・韓国語の各原語の特性に配慮した機能からのアプローチ

##### 1) 中国語同義語検索

一般的に中国語には一つの日本語に対応した複数の中国語表現があり、他の原語に比べると、いわゆる同義語が多いという特性がある。

CNIPRには、原語である中国語と、翻訳された日本語または英語を併用して検索できる、ハイブリッド検索機能がある。一例として「電気自動車」のタイトル検索について、CNIPRのハイブリッド検索機能を使って検証した。

まず、日本語「電気自動車 or 電動自動車」でタイトル検索を行った。次に、一般の辞書サ

イトを使い、「電気自動車」に対応する「电动汽车」「电动车」と、後述のCNIPRの辞書機能を使って抽出した同義語「电动车辆」を併せ、「电动汽车 or 电动车 or 电动车辆」でタイトル検索を行った。その結果、日本語と中国語で検索結果は一致していなかった。ハイブリッド検索機能では、日本語での検索結果と中国語での検索結果を融合することができ、日本語検索のみでヒットした内容を簡単に検討することが可能である。日本語検索のみでヒットしたタイトルを確認した結果、当初の一般辞書サイトと搭載された辞書機能からは抽出できなかった同義語として、「电气车」「电气机动车」「电动机动车」といった他の同義語の存在を把握することができた。当然ながら、実際の特許調査においては、中国語での検索結果と日本語での検索結果の合計和を求めることが可能である。中国語の特性である多数の同義語に対し、日本語で補完検索することによって、ハイブリッド検索機能を駆使したより網羅的な検索の実現が可能となる。

##### 2) 韓国語分かち書き検索

日本語にはない韓国語の特徴として、単語間にスペースを挿入（「分かち書き」と言う）する点あげられる。例えば電気自動車は、「전기자동차」（電気自動車）または「전기 자동차」（電気自動車）のように表現される。韓国語での検索では、分かち書きされたものとそうでないものを区別して検索する必要がある。分かち書きされたものを検索する場合、単語間のスペースを保持したままの検索ができないデータベースでは、「전기 AND 자동차」（電気 AND 自動車）のように二つの単語をAND検索することになる。しかし、この方法では必然的に多くのノイズが発生してしまう。一方、単語間のスペース検索に対応しているデータベースでは、二つの単語を“ ”でくくる、すなわち「“전기 자동차”」（電気自動車）により、スペースを保持したままの検索を行う。

このように、韓国語での検索においては分かち書きに対する留意が必要となるが、データベースによっては、システムの内部で分かち書きを考慮した処理がなされるデータベースがある。一例として、機械翻訳された日本語データを日本語で検索できるデータベースPATBRIDGEでは、「전기 자동차」(電気自動車)はシステムの内部では分かち書きなしで翻訳され、分かち書きなしのまま検索キーとなっている。つまり、分かち書きなしで日本語で検索を行っても、韓国語の分かち書きなし/あり両方の公報を抽出する。従って、韓国語の特性である分かち書きにとらわれることなく、「電気自動車」で一括して検索することが可能である。このようなデータベースの場合、原語由来の特性に精通していないサーチャであっても、機械翻訳された日本語での検索であれば分かち書きを意識する必要がない。

### 3) 韓国語同音異義語検索

韓国語と日本語は文法上の語順がほぼ一致するという特徴があるため、韓国語から日本語への機械翻訳は、他の原語から日本語への翻訳に比べて一般的に精度が高い。しかし一方で、次のような問題点もある。韓国語のハングル表記は表音文字であるため同音異義語が存在し、例えば、「電池」「全地」「前地」は、韓国語でいずれも「전지」と書いて発音も同じになる。このことから、機械翻訳時に誤訳が生じてしまうという問題点がある。韓国語に精通していないと、日本語で検索する際に誤訳が生じる可能性を予測することは困難ではあるが、知っておきたい事項である。そしてこの誤訳という問題点に対しては、次のように誤訳報告機能を備え、対応がなされているデータベースがある。

### (2) 機能メンテナンスによる検索精度の向上からのアプローチ

機能メンテナンスとして、翻訳データのアップ

デートについて検証した。機械翻訳では誤訳といったリスクが伴う。特許データベースを利用して、適当でないと思われる翻訳単語があるために、検索しても正しくヒットしないといった不具合を感じた場合に、サーチャから誤訳報告ができる特許データベース(PATBRIDGE, 中韓翻訳システム)がある。誤訳報告はデータベース作成者により検証され、その結果、誤訳と判断された場合には翻訳辞書が修正される。

しかしここで、修正された翻訳辞書により過去分に遡ってすべてのデータを翻訳し直すかどうかは、各データベースによって異なっている。過去分に遡ることなく、ある時期以降から新しく反映させた辞書を適用するデータベース(中韓翻訳システム)もある。過去分に遡って翻訳が直される場合は、同じ検索語で検索を行っても修正の前後でヒットが異なってくるが、翻訳し直されたことが告知されない限りサーチャが気付くことはないと思われる。ある時期以降から辞書が変更された場合は、その時点から精度が上がることになる。

### (3) 検索キーの展開や拡張といった支援機能からのアプローチ

#### 1) 中国語辞書機能

検索式作成時の補助機能として、翻訳辞書機能を提供しているツールがある。

この翻訳辞書は翻訳エンジンとは別物であることが大半で、現状ではすべての文献をヒットさせるのに十分とはいえない。CNIPRに搭載されている翻訳辞書を使うと「ウェアラブル」の中国語は「服装标签」と表示される。「服装标签」を発明の名称で検索した翻訳を確認すると「胯なり会符」と表示され、翻訳辞書と一致しない。また「ウェアラブル」の中国語は「阿拉拉布」と表示された。しかし、「阿拉拉布」を発明の名称・要約・全請求項で検索しても0件となった。

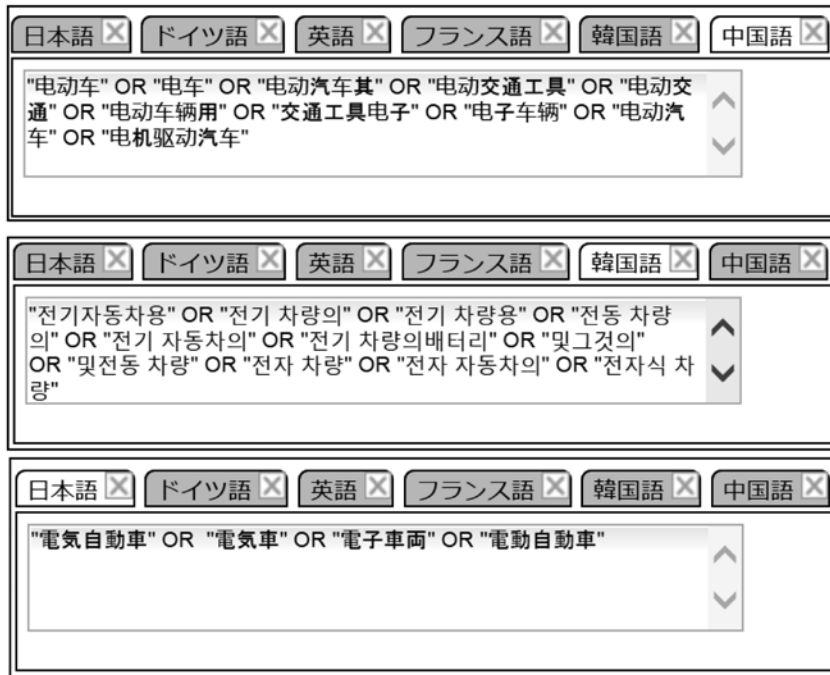


図4 WIPO PATENTSCOPE画面

CNIPRの翻訳辞書は、一部に未対応箇所があるとベンダーから報告を受けている。検索キーワードによっては英語「wearable」に対応する中国語「穿戴」などを併せた検索を行った方が良い結果が得られることが分かった。

### 2) WIPOの多言語検索機能

日本語に対応する中国語、韓国語が分からない場合は、WIPO PATENTSCOPEの多言語検索 (CLIR) を使うことも1つの手段である (図4)。

これに「電気自動車」を入力すると、「電気車」等の類義語が表示される。必要な同義語を追加し、翻訳表示すると中国語では「电动车, 电车, 电动汽车其」などが表示される。タブの切り替えにより韓国語では「전기자동차용, 전기 차량의, 전기 차량용」などが表示されるので、他のツールが提供している翻訳辞書機能より多くの同義語を拾うことができる。

## 3.3 小 括

機械翻訳機能を搭載した中国・韓国特許調査ツールのサーチ機能を検証した結果は、次の通

りである。(1) ハイブリッド検索機能やインターフェースの工夫により、中国語・韓国語の特性を克服、(2) 誤訳に対応した機能のメンテナンスにより検索精度を向上、(3) 辞書機能を充実させることにより、サーチの検索キーの展開や拡張を支援、など、各ツールによって特徴やユーザにメリットとなる独自機能があることが分かった。

目的とする文献を的確にヒットさせるというサーチの立場から、また、そのサーチ環境を整備する管理者の立場からも、これらの機能を搭載したツールの特性を予め把握した上でサーチに適用することが重要である。また、中国・韓国の特許調査においては、ツールの特性の把握のみならず、中国語・韓国語ならではの原語由来の特性を把握した上で、どのツールを選べばよりの確にヒットさせることができるのか、どのように検索すればより原語由来の特性を克服した網羅的な検索が可能になるのか、併せて考慮することも重要である。

## 4. 特許業務に即した評価方法の提言

前章では中国・韓国特許調査ツールのサーチ機能について紹介してきた。当然ながらヒットした特許は中国語であり韓国語である。如何に効率よくサーチ出来たとしても、肝心の内容が理解できなくては目的の文献がヒットしているか否かが不明のままである。サーチャは、まずは機械翻訳ツールを使用し内容の把握に努めることになる。

そこで本章では、中国・韓国特許を機械翻訳による日本語翻訳でスクリーニングする場面を想定し、翻訳ツールについて、その評価方法の検討を行う。

### 4. 1 従来の評価方法に対する意見

冒頭でも触れたように、NMTが実装された翻訳ツールが登場するなど、機械翻訳技術は進歩しているが、自身が普段使用している翻訳ツールとはどう違うのかを明らかにするために、機械翻訳の評価方法に着目した。

機械翻訳の評価方法としては、自動評価の手法として、BLEU, RIBES, TER等が知られている。BLEUは翻訳ツールから出力された翻訳文と参照文とを比較し、一致率を0～1のスコアとして表している。RIBESはBLEUを踏襲しつつさらに語順の正確さにウェイトを置いている。TERは機械翻訳された翻訳文に対して、人

間修正、編集を行った分のコストに着目している。また、人手の評価方法としてはグレード評価があり、例えば「流暢さ」と「正確さ」について人間が5段階で評価している。

しかし、サーチャとしては、知財の業務にどの程度有効なのかが知りたいのであり、スコアやグレードの違いが、実際の知財業務にどのようにマッチするのかが分からない。

そこで、知財業務に即した基準で機械翻訳ツールを評価する必要がある。

### 4. 2 知財業務に即した評価手法

表3に、当小委員会が検討した知財業務に即したカテゴリとその評価例を示す。カテゴリに用いる知財業務は、無効資料調査、クリアランス調査（侵害予防調査）、出願前の先行技術調査、技術動向調査とした。これらは、サーチャがサーチを行う目的として一般的と考えられる業務としている。無効資料調査は権利として確定している特許発明の無効理由に該当する証拠資料調査であり、クリアランス調査は自社製品の侵害予防のための調査であり、先行技術調査は出願前に行う新規性、進歩性の確認のための公知文献調査であり、技術動向調査は特定の技術分野やテーマに関連する特許出願の傾向を調査するものである。

評価を行うにあたり、チェック項目を定め、項目ごとに加点し合計点から評価を行うことと

表3 知財業務に即したカテゴリと評価

知財業務に即したカテゴリ		評価 (請求項の場合)
無効資料調査	無効判断	High
	スクリーニング	High
クリアランス調査	抵触判断	High
	スクリーニング	High
出願前先行技術調査	出願可否判断	High
	スクリーニング	Mid
技術動向調査		Mid



本文の複製、転載、改変、再配布を禁止します。

した。チェック項目は、評価する対象によって項目の内容や点数配分を変える。例えば、明細書の記載個所（請求項はクレームの体裁，要約は主語－述語の繋がり，など）や，異なる技術分野（化学系は化合物名称や数値，機械・構造系は構成要件とその関係や手順，など）毎に定める。即ち評価する対象について，サーチャがより正確な翻訳を期待すると考えられる箇所にウェイトを置いている。

一例として，表4に請求項を評価する場合のチェック項目と配点を示す。

構成要件や名称，手順など，スクリーニングの際に必要な情報をチェック項目として定め，公報の内容を理解するために重要なチェック項目にウェイトを置く配点としている。

また，チェック項目と配点は技術分野ごとに設定すべきと考え，当小委員会では，化学系で

あれば化合物名称や製造工程の手順などを，機械・構造系であれば構成要件の関係／手順をチェック項目として定め，評価点については，機械・構造系は構成要件がより重要と考え，評価点の配分を化学系と変えている。

これらを人手により評価し，その合計点を算出する。ただし，各評価点は中間点の採点をせずに，満点か0点かで評価した。つまり3点の項目であれば，3点か0点のいずれかを付ける。これは，評価が中間点に偏ることを避けたためである。

これら評価点の合計点から次のように総合評価を決定する。Highは満点，Midは3点以上のチェック項目に1つでも0点がある，LowはHigh，Mid以外，である。この総合評価に基づいて，表3を参照しその翻訳ツールが各知財業務に適するかを判定する。例えば，請求項の評

表4 請求項のチェック項目と評価，配点

チェック項目		化学系	機械・構造系
化合物名称	①ハイフン・括弧・数値の位置が正しいか，過不足はないか ②語順が正しいか（IUPAC命名法に沿っているか） ③ギリシャ語表記であるべきところが数値に置き換わっていないか（逆も同様）	3点	－
製造工程の手順	①工程の順番が正しいか	3点	－
構成要件の関係／手順（工程，制御フロー）	①各構成の接続関係が正しいか ②工程・制御フローの順番が正しいか	－	3点
数値	①数値や範囲に間違いはないか ②単位に間違いはないか	3点	－
構成要件	①要件A・B・Cが切り出せているか ②要件Aとその中身 $\alpha$ ，要件Bとその中身 $\beta$ の対応付けが正しいか ③文字化けがあるか ④直訳ではなく正しい技術用語になっているか ⑤複合語が個々の語句の単純な組み合わせになっていないか ⑥同音異義語になっていないか	3点	4点
クレーム体裁	①発明の名称が切り出せているか ②正しい位置（例えば末尾）にあるか ③おいて書きと本文とが区別できているか ④前記・該などが正しく付されているか ⑤従属先の番号が正しいか	1点	1点

表5 評価イメージ

原文（請求項）	要件	クレーム 体裁	数値	化合物 名称	製造 工程 の 手順	合計点	評価
1. 一种高流动耐冲击导电丙烯腈-丁二烯-苯乙烯三元共聚物(ABS)材料, 其特征在于: 所述的材料配方如下(重量百分数): ABS 60~65 炭黑 15~25 抗冲改性剂 15~20 抗氧剂 0.1~0.2 紫外线吸收剂 0.2~0.3 所述的抗冲改性剂为热塑性聚酯弹性体, 或热塑性聚氨酯TPU, 或核-壳结构的接枝物MBS; 所述的配方不含有金属粉末。							
【請求項1】下記組成（質量%）：ABS 60～65 カーボンブラック 15～25 耐衝撃性改良剤 15～20 抗酸化剤 0.1～0.2 紫外線吸収剤 0.2～0.3 を有することを特徴とする高流動性、耐衝撃性および導電性を持つアクリロニトリル・ブタジエン・スチレン三次元共重合体（ABS）組成物。							
導電性のアクリロニトリルに影響を与えるため、高い流動抵抗 - ブタジエン - ことを特徴とするスチレン系共重合体（ABS）材料： 以下の処方（重量パーセント）に記載の材料： ABS60～65 カーボンブラック15～25 耐衝撃性改良剤15から20 酸化防止剤0.1から0.2 0.2から0.3の紫外線吸収剤	0	1	3	3	-	7	Mid
1. 高流動抵抗の耐衝撃性アクリロニトリル - ブタジエン - スチレン三元共重合体（ABS）材料であって、前記材料の配合が以下の通りであることを特徴とする（重量%）。 ABS 60～65 カーボンブラック15～25 インパクトモディファイア15～20 酸化防止剤0.1～0.2 UV吸収剤0.2～0.3	0	1	3	3	-	7	Mid
高流動性耐衝撃性導電性アクリロニトリル-ブタジエン-スチレンターポリマー(abs)材料であって、前記材料式は、下記成分を重量%で含むことを特徴とする： Abs60-65 カーボンブラックは15-25である 15-20部の耐衝撃性改良剤 前記酸化防止剤は、1重量%以上2重量%以下である 紫外線吸収剤は0.2-0.3である	3	1	0	3	-	7	Mid
1.高流動性、耐衝撃性、導電性、アクリロニトリル-ブタジエン-スチレンターポリマー(ABS)材料であり、その特徴は以下のとおりである:材料配合は以下のとおりである(重量百分率)は以下の通りである。ABSは、60～65である。カーボンブラックは、15～25である。耐衝撃性改良剤は、15～20である。酸化防止剤は、0.1～0.2である。紫外線吸収剤は、0.2～0.3である。	3	1	3	3	-	10	High

価がHighの場合、無効資料調査の無効判断に耐えうるレベルと判断する（表5）。

例えばGoogle翻訳（中→日）NMTが無効資料調査の無効判断に使えるかと考えた場合、評価結果「Mid」からは、「無効資料調査での判断にはまだ不十分」であると確認できる。

#### 4.3 評価条件

評価を行うにあたり、評価対象などは下記の条件とした。

翻訳ツールは、中韓翻訳システム、Google翻訳、みんなの自動翻訳<sup>17)</sup>、WIPO Translate、PatentSQUARE、CNIPR（中国）、PATBRIDGE

（韓国）とし、無料・有料、翻訳サイト・検索DBに付属の、各エンジンタイプを織り交ぜた7種類を検討した。

また、技術分野・件数は、中国 7分野：化学系（電池、ポリマー、医薬品、化学製造法、樹脂組成物、樹枝状ポリマー）各1件、機械・構造系（携帯電話）各1件、韓国 3分野：化学系（電池、ポリマー）各1件、機械・構造系（携帯電話）各1件、とした。

対象出願の条件は、優先権の主張国が中国又は韓国、且つ、日本とその他の国（米国、欧州など）に対応出願がある、グローバルでビジネスを行っている企業からの出願のものとし、

Derwent Innovation<sup>18)</sup>により抽出した。このような企業であれば、中国語・韓国語は元より、日本語としても正しさが必要であり、出願数の多さからも原文、対応出願の日本語訳共に文章の質が一定の水準を満たしていると期待できる為である。また、上記とともに、当委員会で2012年度に報告している論説<sup>7)</sup>記載のものについても再検討した。

正解とする日本語訳は、対応日本出願の日本語訳（また場合により英訳を参考）とした。

技術分野別に準備したチェック項目のパターンは、化学系、機械・構造系の2パターンとし、翻訳対象箇所は請求項、要約の2箇所とした。

翻訳ルートは、中・韓→日、中・韓→英語→日の2ルートとした。

#### 4. 4 評価結果に対する考察

次に評価結果について考察する。

表6、表7に示すように、中国特許翻訳ツールの評価結果のばらつきは大きく、現時点では、翻訳ツールはこれが良いと評価できる段階では

ないと考える。また同じ翻訳ツールであっても案件により評価結果にばらつきがあった。

一方、中国特許翻訳に比べ、韓国特許翻訳は高評価であった。3. 2 (1) 3) 韓国語検索の検証例でもふれた通り日本語と韓国語の語順や文法がほぼ同じで、機械翻訳が比較的容易であるためと考えられる。

また、技術分野による傾向はみうけられなかった。但し、同じ案件の過去の翻訳結果<sup>6), 7)</sup>と比較すると、確実に進化していると言える。

次に、同じ翻訳ツールであっても案件により評価結果がばらつく事象について具体例を挙げて説明する。表8に示すように、例えば、みんなの自動翻訳（NMT）の場合、一部重複箇所があるものの、翻訳内容はほぼ正解文といえる精度であった一方、化合物の名称が全く異なるものとなる場合があった。

以上のことから、ツールをサーチャにも解るように知財業務に即した表現でカテゴリ化したことにより、サーチャは目的の知財業務に活用できるか確認可能になる。また、技術分野に

表6 翻訳ツールの評価結果

翻訳システム	エンジンタイプ*	結果 化学系・請求項
中国特許翻訳		
中韓翻訳システム	RBMT	High~Low
Google翻訳	NMT	High~Low
みんなの自動翻訳	NMT	High~Low
WIPO Translate	NMT	High~Low
CNIPR	RBMT or SMT	High~Low
PatentSQUARE	RBMT or SMT	High~Low
韓国特許翻訳		
中韓翻訳システム	RBMT	High
Google翻訳	SMT	High
みんなの自動翻訳	SMT	High
WIPO Translate	NMT	Mid~Low
PATBRIDGE	RBMT or SMT	High
PatentSQUARE	RBMT or SMT	Low

\*エンジンタイプは、Google翻訳、みんなの自動翻訳、WIPO TranslateはNMTと公表、その他は独自の調査結果による。

表7 中国特許 請求項の評価結果

番号 2012年度案件 技術分野	CN1691373A 電池 2012年度 化学系	CN1282762A 化合物(ポリマー) 2012年度 化学系	CN101245050A 化合物(医薬品) 2012年度 化学系	CN104448702A 樹脂組成物 化学系	CN104558628A 樹枝状ポリマー 化学系	CN104557437A 化学製造法 製造系	CN1337812A 携帯電話 2012年度 機械・構造系
Google翻訳 (中→日)	Mid	Mid	High	Low	Mid	Low	Mid
Google翻訳 (中→英→日)	Mid	Mid	High	Low	Mid	Low	Mid
みんなの自動翻訳 JPO特許【中国語(簡体字) - 日本語】	Mid	High	Low	Low	Mid	Mid	Low
WIPO Translate (中→英→日) Domain:[automatic detection]	Low	Mid	Mid	Low	Mid	Low	Mid
CNIPR	Low	High	High	Low	Mid	Mid	Low
PatentSQUARE	Low	Mid	Low	High	High	High	Mid
中韓翻訳システム	Low (参考:登録公報)	High (参考:登録公報)	Low	High	High	Low	High (参考:登録公報)

表8 同じ翻訳ツールであっても案件により評価結果がばらついた例

良好な翻訳例 ; CN1282762A 化合物(ポリマー)

対応日本特許 特表2003-529626	【請求項1】下記組成(質量%) : ABS60~65 カーボンブラック15~25 耐衝撃性改良剤15~20 抗酸化剤0.1~0.2 紫外線吸収剤0.2~0.3を有することを特徴とする高流動性、耐衝撃性および導電性を持つアクリロニトリル・ブタジエン・スチレン三次元共重合体(ABS)組成物。
みんなの自動翻訳 (JPO特許【中国語(簡体字) - 日本語】)	1.高流動性、耐衝撃性、導電性、アクリロニトリル-ブタジエン-スチレンターポリマー(ABS)材料であり、その特徴は以下のとおりである:材料配合は以下のとおりである(重量百分率)は以下の通りである。ABSは、60~65である。カーボンブラックは、15~25である。耐衝撃性改良剤は、15~20である。酸化防止剤は、0.1~0.2である。紫外線吸収剤は、0.2~0.3である。

一部重複があるものの、翻訳内容はほぼ正解文

不良な翻訳例 ; CN101245050A 化合物(医薬品)

原文に適したと思われる訳請求項7	塩は、テトラメチルアンモニウム塩、テトラエチルアンモニウム塩、メチルアミン塩、ジメチルアミン塩、トリエチルアミン塩である請求項6に記載の塩。
みんなの自動翻訳 (JPO特許【中国語(簡体字) - 日本語】)	7.前記塩が4-メチル-アミン、4-エチルアミン塩、メチルアミン塩、ジメチルアミン塩、トリエチルアミン塩、エチルアミン塩、請求項6に記載の塩。

化合物の名称がまったく異なるものになっている

より重要度に違いがあるため技術分野によりチェック項目を変える必要があることが分かった。但し、各翻訳ツールの評価結果からは、どの翻訳ツールが良いと評価できる段階ではないと考える。

現状では、これら翻訳ツールを知財業務にどのように使い分ければよいかという点について、誤訳エラー事例から考察する。

2章で述べたように、RBMTでは翻訳の質が

安定しており、訳抜けや漏れはないが新語への対応が必要である。一方、NMTは読みやすい翻訳文を生成するものの、訳抜けや過翻訳が生じる。

表9に示すように、中韓翻訳システムなどのRBMTは訳抜けや漏れはないが、化合物名等の翻訳で、原文のままになるものが見受けられた。

Google翻訳での韓国、以前の中国などのSMTでは、適した例文が存在していた場合は



表9 翻訳エラーの事例

RBMT (ルールベース機械翻訳)

<p>CN104558628A 要約 原文に適したと思われる訳</p>	<p>本発明は磁性ナノスケールドンドリマー化合物およびその調製方法を開示する。その分子式Iに示すようである：<math>\Gamma (CH_2)_3 N_{(2^{n+1}-1)} R^1_{(2^{n+2}-2)} R^2_{(2^{n+1})}</math> I。そのうち、<math>\Gamma</math>は磁性ナノ粒子を表し、前記<math>(CH_2)_3 N_{(2^{n+1}-1)} R^1_{(2^{n+2}-2)} R^2_{(2^{n+1})}</math>は樹枝状構造を有する基であり、前記<math>R^2</math>〔注：本来は<math>R^2_{(2^{n+1})}</math>で原文の間違ひと思われる〕は親油性ラジカルでかつ、<math>0 \leq n \leq 100</math>。さらに前記磁性ナノスケールドンドリマー化合物を含む潤滑剤を公開する。</p>
<p>中韓翻訳システム 要約</p>	<p>本発明は磁性ナノ級木枝分子量化合物およびその調製方法を開示して、その分子式Iに示すように：<math>\Gamma (CH_2)_3 N_{(2^{n+1}-1)} R^1_{(2^{n+2}-2)} R^2_{(2^{n+1})}</math> Iはそのうち、<math>\Gamma</math>は磁性ナノ粒子を表して、前記<math>(CH_2)_3 N_{(2^{n+1}-1)} R^1_{(2^{n+2}-2)} R^2_{(2^{n+1})}</math>は樹枝状構造を有するラジカルとし、前記<math>R^2</math>は親油性ラジカルとして、<math>0 \leq n \leq 100</math>。本発明では前記磁性ナノ級木枝分子量化合物を含む潤滑剤を公開する。</p>

化合物名 (木枝=デンドリマー) が訳せていない

NMT (ニューラル機械翻訳)

<p>CN104558628A 要約 原文に適したと思われる訳</p>	<p>本発明は磁性ナノスケールドンドリマー化合物およびその調製方法を開示する。その分子式Iに示すようである：<math>\Gamma (CH_2)_3 N_{(2^{n+1}-1)} R^1_{(2^{n+2}-2)} R^2_{(2^{n+1})}</math> I。そのうち、<math>\Gamma</math>は磁性ナノ粒子を表し、前記<math>(CH_2)_3 N_{(2^{n+1}-1)} R^1_{(2^{n+2}-2)} R^2_{(2^{n+1})}</math>は樹枝状構造を有する基であり、前記<math>R^2</math>は親油性ラジカルでかつ、<math>0 \leq n \leq 100</math>。さらに前記磁性ナノスケールドンドリマー化合物を含む潤滑剤を公開する。</p>
<p>Google翻訳 要約 (中→日)</p>	<p>本発明は、磁性ナノデンドリマー化合物およびその調製方法を開示し、分子の式は式Iに示されるとおりである：<math>\Gamma (CH_2)_3 N_{(2n+1-1)} R^1_{(CH_2)_3 N_{(2n+1-1)}}</math> R1は樹枝状構造を有する基であり、前記R2は<math>0 \leq n \leq 100</math>である親油性基である。本発明は、磁性ナノスケールドンドリマー化合物を含有する潤滑剤を開示する。</p>

$\Gamma_{(2^{n+2}-2)} R^2_{(2^{n+1})}$  I。そのうち、 $\Gamma$ は磁性ナノ粒子を表し、 $\Gamma$ が欠落

良好な翻訳となるが、符号や化合物表記が存在すると、不正確になることが見られた。

Google翻訳、みんなの自動翻訳、WIPO TranslateなどのNMTは、読みやすい翻訳文を生成するが、重要な記述が抜け、違った数値が入るものが見られた。

このことから、エンジンタイプ毎に得意/不得意があり、RBMT (中韓翻訳システムなど)とNMT (Google翻訳など)を組み合わせると互いの欠点を補えることが分かり、業務ではそれぞれの特徴を知った上で併用し結果を確認しながら使うべきといえる。

現在、英語特許に関しては、SMTとNMTを組み合わせた翻訳“ハイブリッド翻訳”を備えたツールが実用化されており、中国語特許についても開発が進められている。今後の翻訳技術の進化に期待する。

## 5. まとめと提言

本稿では、中国・韓国特許を調査するケースを取り上げ、外国特許調査と機械翻訳技術を概

観した。原文と翻訳文をサーチするツール、スクリーニングでの翻訳レベルと、特許調査業務のマッチングについて、検証、考察した。

### (1) 外国特許調査と機械翻訳

特許DBの蓄積翻訳文へのサーチ、Webサービスのリアルタイム翻訳など、外国特許調査での機械翻訳技術を利用する場面は広がってきている。Google翻訳の検証では、AI技術により出力される翻訳文において、以前からの飛躍的な精度向上が確認できた。またGoogle翻訳は本研究期間内でも出力される翻訳文が変化する状況で、日々改良がなされている様子である。機械翻訳システムは今後もアルゴリズムの改良や学習による精度向上が見込まれる。

### (2) 中国・韓国特許のサーチ機能

外国特許のサーチ環境として、中国・韓国の特許調査ツールについて検証した結果、各原語の特性に配慮した機能や、原語に明るくないサーチャに配慮した支援機能があることが分かつ

た。様々な機能を搭載したツールの特性を予め把握した上でサーチに適用すべきであるのはもちろんのこと、今回の検証では、中国語・韓国語ならではの原語由来の特性を把握することも重要であることが分かった。上記を、ともに考慮することが有効であると考え。また、その時点における状況だけでなく、今後の原文表記の移り変わりや、それらに対する各データベースの機能改良状況を常にウォッチし、我々自身に対応していく必要があるのは言うまでもない。

### (3) スクリーニングでの翻訳精度評価

翻訳ツールをサーチャにも解るように知財業務に即した表現でカテゴライズすることにより、目的の知財業務に活用できるかが確認可能になる。また、技術分野により重要度に違いがあるため技術分野によりチェック項目を変える必要があることが分かった。但し、現時点では、どの翻訳ツールが良いと評価できる段階ではないと考える。

また、機械翻訳のエンジンタイプにより得意／不得意はあるものの、違うタイプの翻訳エンジンを組み合わせると互いの欠点を補えることが分かった。業務ではそれぞれの特徴を知った上で併用し結果を確認しながら使うべきであると考え。

## 6. おわりに

機械翻訳技術は、外国特許調査業務の効率化の点でますます有効な存在になってきた。その理由としては大きく3つある。1つ目は、特許分野への適用が進み特許データベースやWeb翻訳サービスなどツールの選択肢が増加していることである。特に無料でサービス提供されるツールも増え、企業が利用するハードルが低くなった。2つ目は、これまでの「逐語訳的ではあるが内容検討に向けた正確な翻訳」に加え、「流暢で読みやすい翻訳」が利用可能になり、特許

業務での活用場面が広がったことである。これにより特定の専任者以外でも同等の業務を行うことが可能になり、対応できる人員の幅が広がることで、知財部員リソース配分の自由度が上がる。3つ目は、「翻訳に起因する調査リスク」を知ることでツールの弱点を相互にカバーできるようになったことである。ツールの調査リスクを知って「利便性と業務とのマッチング」させることで、業務の精度を確保しながら効率化できると考える。

したがって、ツールで使用される翻訳エンジンの種類や特徴、ツールの検索機能の設計ポリシーや翻訳文データの提供ポリシーなどを把握することは、自社の外国特許調査業務に適切なツール選択をする有効なヒントとなる。単に、新しいツールあるいは新しい翻訳エンジンというだけの理由ですぐに採用したり、また、翻訳による調査リスクを確認することもなく1種類のツールを使い続けてしまったり、という行動は避けるべきであろう。機械翻訳技術およびツールは日々改良されている。今後もツールの翻訳精度状況の変化を把握し対応することが必要である。

機械翻訳技術を利用する外国特許調査業務の効率化について、本稿が会員企業での知財調査活動業務の一助となれば幸いである。

本稿は、2017年度情報検索委員会第1小委員会委員、加藤昌央（パナソニック、小委員長）、風間進二（旭化成）、金子裕二（カルソニックカンセイ）、柴田潔子（住友電工知財テクノセンター）、鈴木順行（三井化学）、土居美緒（第一三共）、東村恵理（DIC）の執筆によるものである。

### 注 記

- 1) WIPO Develops Cutting-Edge Translation Tool For Patent Documents  
Geneva, October 31, 2016.

- [http://www.wipo.int/pressroom/en/articles/2016/article\\_0014.html](http://www.wipo.int/pressroom/en/articles/2016/article_0014.html)
- 2) Found in translation: More accurate, fluent sentences in Google Translate  
Nov 15, 2016.  
<https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/>
  - 3) 田畑, ニューラル翻訳を用いた中国特許機械翻訳精度の検証, 第14回情報プロフェッショナルシンポジウム予稿集, A32, 2017.
  - 4) 中澤, 機械翻訳の新しいパラダイム ニューラル機械翻訳の原理, 情報管理. 2017, vol.60, no.5, pp.299-306.
  - 5) 知的財産情報検索委員会第3小委員会 2005年度, 知財管理, Vol.56, No.12, pp.1881-1893 (2006).
  - 6) 知的財産情報検索委員会第2小委員会 2012年度, 知財管理, Vol.63, No.11, pp.1763-1776 (2013).
  - 7) 知的財産情報検索委員会第2小委員会 2012年度, 知財管理, Vol.63, No.12, pp.1943-1957 (2013).
  - 8) CNIPR日本版, 株式会社プロパティ.
  - 9) PATBRIDGE, 韓国WIPS社.
  - 10) 中韓文献 翻訳・検索システム, 特許庁.
  - 11) PatentSQUARE, パナソニックソリューションテクノロジー株式会社.
  - 12) EPO worldwide legal status database (INPADOC),  
<https://www.epo.org/searching-for-patents/legal/inpadoc.html>
  - 13) Derwent World Patents Index (DWPI),  
<https://clarivate.jp/products/derwent-world-patents-index/>
  - 14) Espacenet Patent search, EPO,  
[https://worldwide.espacenet.com/advancedSearch?locale=jp\\_JP](https://worldwide.espacenet.com/advancedSearch?locale=jp_JP)
  - 15) PATENTSCOPE国際・国内特許データベース検索, WIPO,  
<https://patentscope.wipo.int/search/ja/structuredSearch.jsf>
  - 16) 国立研究開発法人情報通信研究機構プレスリリース (2017.6.28), 「ニューラル機械翻訳で音声翻訳アプリVoiceTraが更なる高精度化を実現」,  
<https://www.nict.go.jp/press/2017/06/28-1.html>
  - 17) みんなの自動翻訳, 国立研究開発法人情報通信研究機構 (NICT).
  - 18) Derwent Innovation,  
Clarivate Analytics (Japan) Co., Ltd.,  
<https://clarivate.jp/products/derwent-innovation/>
- (URLの参照日は全て2018年7月23日)

#### 参考文献

- ・中国・韓国語の特許文献を日本語で検索可能なシステムのご紹介, 櫻井健太, tokugikon 2015.1.28 No.276 pp.67-70.
- ・中韓文献翻訳・検索システム: 検索漏れの低減案: 収録率及び翻訳不良の文法的な観点からの検証, 西尾潤, 雙田飛鳥, 石田政司, 伊藤徹男  
情報プロフェッショナルシンポジウム予稿集 Vol.12th, pp.165-170.

(原稿受領日 2018年9月27日)