

# 特許庁における特許情報の翻訳に係る取組み —機械翻訳プラットフォームの活用等—

Approaches to translating patent information at the JPO



特許庁 総務部総務課特許情報室特許情報利用推進班長

中西 聡

平成 17 年特許庁入庁。化学分野の特許審査や特許権存続期間の延長登録の審査に従事のほか、調整課審査基準室、審査推進室を経て平成 31 年 1 月より現職。

## 1 はじめに

国内出願人による特許及び実用新案の出願件数<sup>1</sup>を出願先別にみると、国内への出願が減少傾向にある中、海外への出願は、増加ないし横ばいの傾向にあって、その割合が年々増加している。このことから、国内出願人の、海外への出願に対する関心や重要度の高さが窺える。

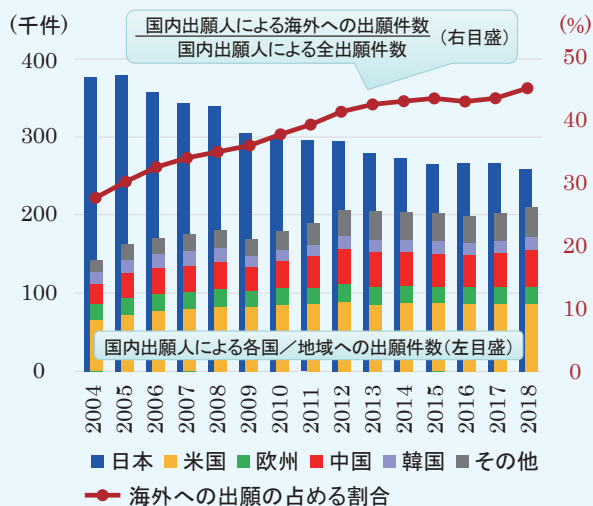


図 1 国内出願人による出願先別の出願件数 (特実) (PCT の国内移行件数を含む)

また、世界各国又は地域への特許及び実用新案の出願件数をみると、五庁 (日米欧中韓) への出願が 8～9 割を占める中、海外、特に中国への出願が急増している。

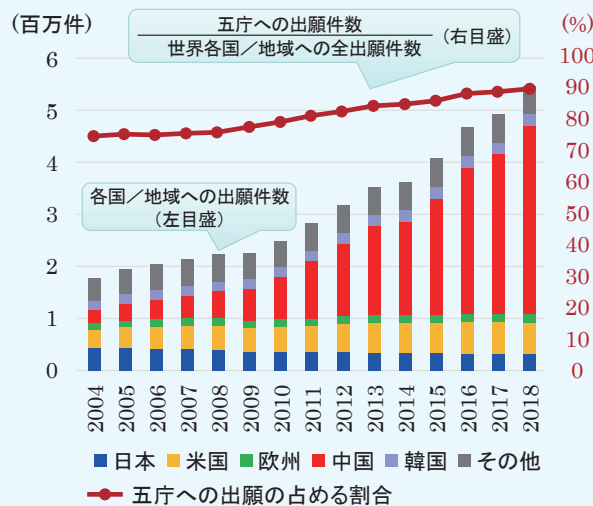


図 2 世界各国又は地域への出願件数 (特実) (PCT の国内移行件数を含む)

これら背景の下、特許庁では、国内出願人の関心や重要度の高さが窺える海外への出願において、国内出願人が円滑に権利取得するための一助となるように、国内の特許情報を海外発信する取組みや、海外への出願、特に、急増している中国への出願の検索、照会環境を整備する取組みを実施している。

いずれの取組みにおいても、日本語以外の言語を扱うことから、翻訳が重要な位置づけにあるところ、特許庁では、機械翻訳プラットフォーム (MTP: Machine Translation Platform) を利用して、国内の特許等に係る各種公報及び審査書類等の英訳、並びに、中国及び韓国の特許及び実用新案に係る各種公報の和訳を実施している。

そして、MTP によって作成された英訳文又は和訳文

1 筆頭出願人の住所又は居所等が国内である出願の件数

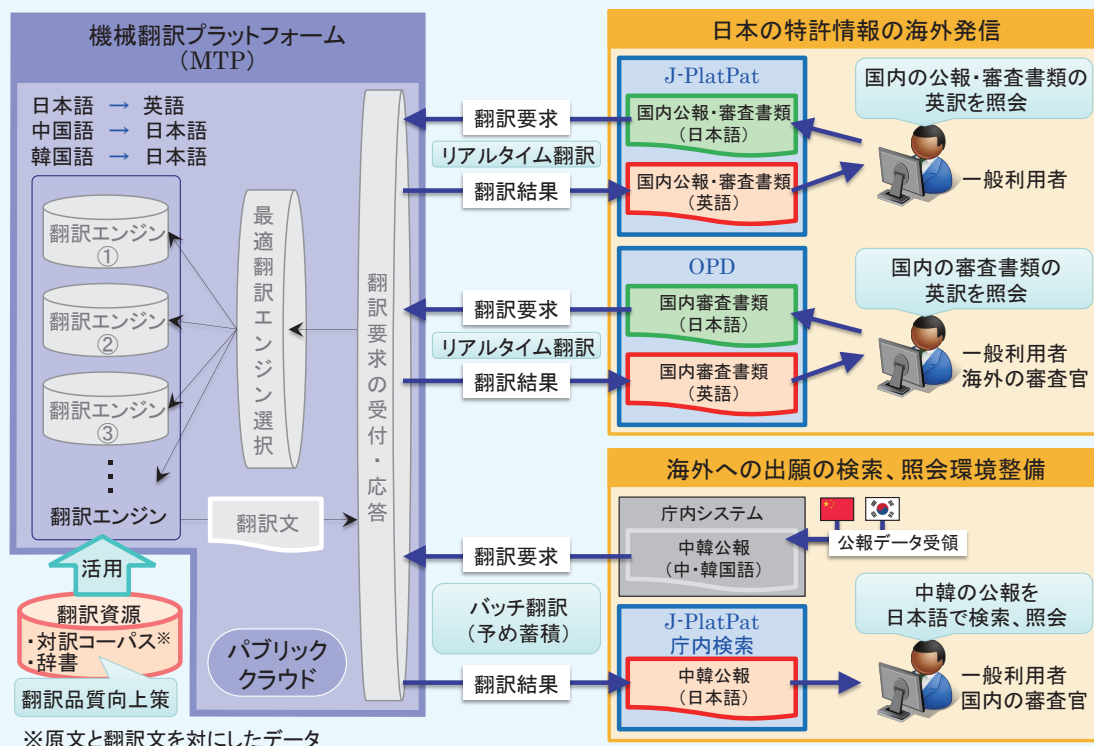


図3 機械翻訳プラットフォーム (MTP)

は、特許情報プラットフォーム (J-PlatPat)、ワン・ポータル・ドシエ (OPD) 又は特許庁内の検索システム等を通じて、一般利用者や国内外の審査官等に向けて無料で提供されており、前述した国内の特許情報の海外発信や、海外への出願の検索、照会環境整備に寄与している。

本稿では、現在運用中の MTP に関する話題を中心に、特許庁における特許情報の翻訳に係る取組みについて述べる。

## 2 機械翻訳プラットフォーム (MTP)

### 2.1 MTP の概要

特許庁は、平成 30 年 4 月に、MTP の設計・開発及び運用サービス一式の契約を東芝デジタルソリューションズ株式会社と締結した。構築された MTP は、令和元年 5 月の日英翻訳機能リリース、及び、翌令和 2 年 4 月の中日・韓日翻訳機能リリースを経て、現在運用中である。

なお、MTP のリリースを理由として、東芝デジタルソリューションズ株式会社は、AAMT 長尾賞<sup>2</sup>の第 15 回 (2020 年) 受賞者に選定された。

MTP は、パブリッククラウドに構築されており、複数

種類の翻訳エンジンを搭載している。そして、J-PlatPat 等を通じた各種公報等の翻訳要求を受け付けると、翻訳対象に応じた最適な翻訳エンジンの選択及び翻訳等を実施し、作成した翻訳文を応答する仕組みになっている。また、翻訳資源を活用した、各翻訳エンジンの翻訳品質向上策が随時実施できるようになっている。

国内の公報又は審査書類等の英訳文は、J-PlatPat 又は OPD 上で照会される度に、MTP によって作成される (リアルタイム翻訳)。対して、中韓公報の和訳文は、中韓から受領した公報データを元に、MTP によって定期的に作成され、検索システムに予め蓄積されている (バッチ翻訳)。

リアルタイム翻訳では、翻訳文が蓄積されておらず、翻訳文を対象としたテキスト検索はできないが、その代わりに、照会時点までに実施された翻訳品質向上策が全て反映された翻訳文が照会できる。対して、バッチ翻訳では、翻訳文が予め蓄積されており、翻訳文を対象としたテキスト検索ができるが、その代わりに、照会できる翻訳文は、過去の蓄積時点で作成されたものであって、その翻訳品質がリアルタイム翻訳に劣る場合がある。

公的なサービスが提供する機械翻訳については、当面、翻訳精度の向上が期待されることから、リアルタイム翻訳を原則とし、例外として、既に中韓文献翻訳・検索シ

2 AAMT 長尾賞 <https://aamt.info/news/nagao-2/>

ステム<sup>3</sup>（J-PlatPat への移行に伴って令和元年6月に終了）においてバッチ翻訳を開始していた中韓公報の機会翻訳は、バッチ翻訳を継続するという方針が示されている<sup>4</sup>。MTP における日英のリアルタイム翻訳と中日・韓日のバッチ翻訳の別は、この方針に沿うものとなっている。

## 2.2 高い翻訳品質

MTP では、ニューラル機械翻訳（NMT: Neural machine translation）がメインに利用されており、NMT 以外の翻訳エンジンである統計的機械翻訳（SMT: Statistical machine translation）及びルールベース機械翻訳（RBMT: Rule-based machine translation）の併用や、翻訳前後の文章に対する各種処理等を含む独自システム<sup>5</sup>を通じて、高い翻訳品質を達成している（3.2 で述べる BLEU の平均スコアとして、国内公報の明細書の英訳文は 40 以上、中国及び韓国公報の明細書の和訳文は 45 及び 60 以上）。

ただし、中韓公報の和訳には、2.1 で述べたとおりバッチ翻訳が採用されているところ、J-PlatPat 等を通じて検索、照会ができる中韓公報の和訳文のうち、MTP によって作成された高品質なものは、概ね令和2年3月の発行分からである。それ以前の発行分は、2.1 で述べた中韓文献翻訳・検索システムにおいて和訳されたものである。

## 2.3 NMT の活用

### 2.3.1 ニューラルネットワーク

MTP においてメインに利用されている NMT は、ニューラルネットワークを活用した翻訳エンジンであって、AI 翻訳とも呼称される。

ニューラルネットワークは、人間の脳神経ネットワークを模して構成されており、脳神経（ニューロン）に見立てた大量の数式（人工ニューロン）が相互に連結した

数式の集合体といえる。入力データは数値等に変換され、前記数式によって順次計算された結果が出力データに変換される。

ニューラルネットワークの出力データを妥当なものとするためには、上記計算結果が適切な値になるように、各数式のパラメータを最適化する必要がある。そして、この最適化は、大量の学習用データ（例えば、入力データと正解の出力データとの対）を使用した学習（ディープラーニング）によって行われる。

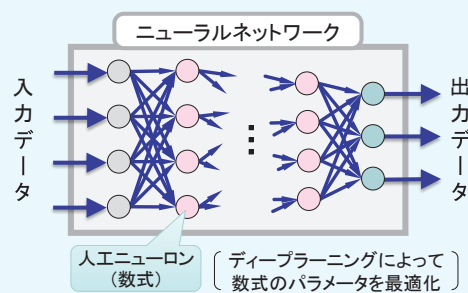


図4 ニューラルネットワークの模式図（注<sup>6</sup>）

### 2.3.2 対訳コーパス

NMT における学習用データとしては、原文（入力データ）と訳文（正解の出力データ）を対にした、対訳コーパスと呼称されるものが主に使用される。

そして、NMT のニューラルネットワークは、実際の翻訳対象の文章と同様の文章に由来する対訳コーパスを学習すればするほど、翻訳対象に特化した最適化が進行し、翻訳品質の向上が期待できるようになる。一方、訳文に省略や追記等の不備がある対訳コーパスを学習してしまうと、省略に起因した訳抜け（原文を部分的に翻訳しない誤訳）や、追記に起因した湧き出し（原文に無い用語を訳文に追加する誤訳）の発生原因になる。

したがって、翻訳品質に優れた NMT を構築するためには、実際の翻訳対象の文章と同様の文章に由来する対訳コーパスであって、その訳文に不備がない（又は少ない）優良なものを大量に学習させる必要がある。

通常、このような優良な対訳コーパスの作成には、人手翻訳が必要となることから、大量に作成するためには多くの費用と時間がかかる。しかし、特許等の公報については、パテントファミリーの公報を活用して、比較的優良な対訳コーパスを大量に機械的に作成することがで

3 檀本英吾、中韓文献翻訳・検索システム、Japio YEAR BOOK 2014、p.62-65

4 情報普及活用小委員会報告書「特許情報のさらなる活用に向けて」（p.31 脚柱 16 参照）<https://www.jpo.go.jp/resources/shingikai/sangyo-kouzou/shousai/jouhouhukyu-shoi/h28houkokusho.html>

5 園尾聡、ニューラル機械翻訳による特許機械翻訳システムの開発、Japio YEAR BOOK 2019、p.296-301

6 NMT のニューラルネットワークは、人工ニューロンの配置や連結態様等が工夫されており、模式図どおりではない。

きる。その理由として、(1) 翻訳対象の文章と同様の文章によって構成されている点、(2) 各国又は地域に同内容の出願をするという目的から、パテントファミリーの公報の文章同士は、省略や追記等の相違が少ない対訳の関係にある場合が多い点、(3) 公開情報として大量に存在している点、等が挙げられる。

そこで、MTPにおけるNMTの構築にあたっては、パテントファミリーの公報に由来する対訳コーパスが最も多く活用された。また、MTPの翻訳対象には公報以外も含まれるところ、その対応等のためにパテントファミリーの公報以外に由来する対訳コーパスも活用された。

主なものを挙げると、日本語／英語の対訳コーパスとして、特許庁と国立研究開発法人情報通信研究機構(NICT)の協力の下に作成した、パテントファミリーの公報に由来する数億規模のもの、及び、特許庁で実施した調査事業<sup>7</sup>等を通じて作成した、国内の各種審決、判決又は審査書類等に由来する十数万規模のものが挙げられる。また、中国語／日本語の対訳コーパスとして、特許庁で実施した調査事業<sup>8</sup>等を通じて作成した、パテントファミリーの公報等に由来する数億規模のものが挙げられる。

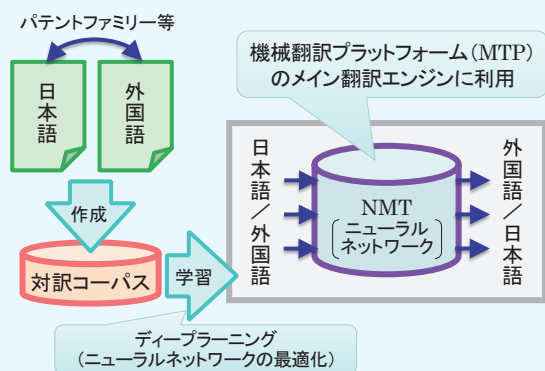


図5 対訳コーパスの作成・学習

なお、韓日翻訳は、SMTによって十分な翻訳品質が達成できることから、MTPにおいてはSMTがメインに利用されている。そして、詳細は専門書に譲るが、SMTの構築にも対訳コーパスが活用されるどころ、MTPにおけるSMTの構築に活用した韓国語／日本語

の対訳コーパスとして、特許庁とNICTの協力の下に作成した、パテントファミリーの公報に由来する1億弱規模のものが挙げられる。

### 2.3.3 辞書

ニューラルネットワークでは、入力データ全体が一度に計算対象となって出力データが作成されることから、NMTに入力された翻訳対象文は、用語単位で翻訳されるのではなく、その文全体が一度に翻訳対象となって翻訳される。したがって、同じ用語であっても、その用語を含む翻訳対象文が異なれば、NMTの翻訳結果は異なる場合がある。

しかしそれでは、一貫して同じ翻訳語に翻訳されることが好ましい定型の語句(例:請求項/claim)の翻訳にゆれが生じる。そこでMTPにおけるNMTには、この翻訳ゆれを防止するために、辞書機能が組み込まれている。この機能によって、辞書に登録した用語であれば、原則として辞書のとおりに翻訳されるようになっていく。

特許庁では、各種調査事業<sup>9</sup>等を通じて、各種公報や審査書類等に由来する辞書を作成し、例えば日英翻訳について約70万語をMTPの辞書に登録した。

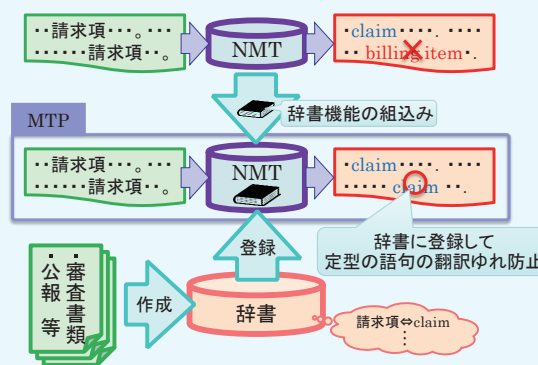


図6 辞書の作成・登録

なお、MTPでは、NMT以外のSMT及びRBMTの翻訳エンジンにも辞書機能が組み込まれており、定型の語句の翻訳に利用されている。

## 3 翻訳エンジンの性能向上の取組み

### 3.1 NICTとの連携・協力

特許庁は、機械翻訳の翻訳品質向上や活用促進等を

7 平成27年度から平成30年度に実施した、審決の日英機械翻訳に関する調査等の各種調査事業

8 平成24年度から平成29年度に実施した、中国特許文献の機械翻訳のための中日辞書整備及び機械翻訳性能向上に関する調査等の各種調査事業

9 脚柱7、8の調査事業

目的として、情報通信分野を専門とする公的研究機関の国立研究開発法人情報通信研究機構（NICT）と、平成26年7月から連携、協力関係にある。

特許庁とNICTとは、協力して様々な取組みを実施している。MTPに関連する例を挙げると、MTPに搭載しているNMT及びSMTの翻訳エンジンは、NICTで開発されたものであるし、2.3.2で述べたように、協力して対訳コーパスを作成している。

また、NICTにおいて開発された翻訳エンジンは、MTPだけでなく、別途、特許庁の審査官向けにも活用している。例えば、特許庁では、自由に入力した英語等の外国語をリアルタイムで和訳することができる、多言語の機械翻訳環境が庁内の審査官向けに展開されているところ、この翻訳エンジンに活用している。

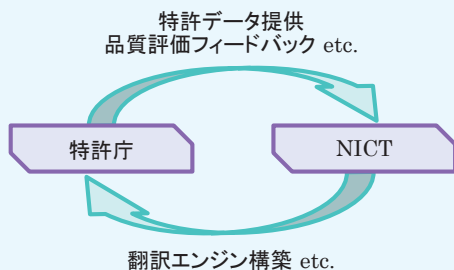


図7 特許庁とNICTとの連携・協力

### 3.2 翻訳品質の評価

翻訳品質向上策の効果確認や、各種翻訳エンジンの相互比較等のために、翻訳エンジンの性能を客観的に評価する方法が必要であるところ、その一つとして、作成された機械翻訳文の翻訳品質を評価する方法が挙げられる。

翻訳品質を評価する方法は、人手評価と自動評価の2つに大きく分けられる。

人手評価とは、人間が原文と機械翻訳文を見比べて翻訳品質を評価する方法である。特許庁では、特許情報の機械翻訳文に対する翻訳品質の適切な人手評価方法として、原文と、別途用意した参照用の正解翻訳文と、評価対象の機械翻訳文とを元にスコア化する方法を公表している<sup>10</sup>。

自動評価とは、機械が自動的に翻訳品質を評価する方法である。著名なものの一つとしてBLEU（Bilingual

Evaluation Understudy<sup>11</sup>）スコアを利用する方法がある。BLEUスコアは、評価対象の機械翻訳文と、別途用意した参照用の正解翻訳文との類似度を、所定の計算式によって0から1の間の値に数値化したものである。通常はパーセント（100倍した値）で表現し、値が大きい（類似度が高い）ほど品質が高いとみなされ、一般的には、スコアが40以上であると高品質と解される。

なお、MTPの翻訳品質は、上記公表した方法による人手評価と、BLEUスコア等を利用した自動評価とを併用して管理している。

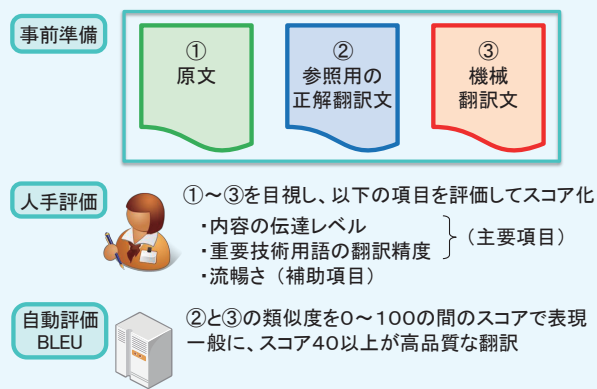


図8 機械翻訳文の翻訳品質の評価

### 3.3 対訳コーパス・辞書の作成

MTPの翻訳エンジンの構築にあたって活用された対訳コーパス・辞書の作成について2.3.2及び2.3.3で述べたが、特許庁では、現在も、NMTの性能向上等に資する翻訳資源として、対訳コーパス・辞書を作成する事業<sup>12</sup>を実施している。

これら事業を通じて、令和3年度末までに、特許の審査書類（拒絶理由通知書等）に由来する数百万規模の対訳コーパス及び数万規模の辞書（日本語／英語）、並びに、中国の審決に由来する数百万規模の対訳コーパス及び十万規模の辞書（中国語／日本語）を作成する予定である。

上記対訳コーパス及び辞書は人手翻訳を通じて作成されることから、特に対訳コーパスは、訳抜けや湧き出しの発生原因となる翻訳文の不備（2.3.2参照）がなく、

11 PAPINENI, K., et al., BLEU: a Method for Automatic Evaluation of Machine Translation, Proc. ACL 2002, 311-318

12 審査書類・審決の機械翻訳に関する解析及びデータ作成事業、及び、中国審決情報の機械翻訳に関する解析及びデータ作成事業

10 特許文献機械翻訳の品質評価手順について [https://www.jpo.go.jp/system/laws/sesaku/kikaihonyaku/tokkyohonyaku\\_hyouka.html](https://www.jpo.go.jp/system/laws/sesaku/kikaihonyaku/tokkyohonyaku_hyouka.html)

NMT のディープラーニングに有用と考えられ、今後の NMT の性能向上等に大いに役立つことが期待される。

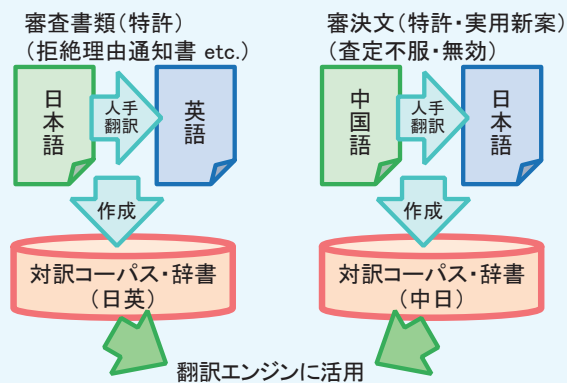


図9 対訳コーパス・辞書の作成

### 3.4 未知語対応

ニューラルネットワークは、学習用データを基に最適化される(2.3.1 参照)から、学習用データに存在しない入力に対する正しい出力は期待できない。したがって、NMT では、学習用の対訳コーパスに存在しない用語(未知語)があると、誤訳となる場合がある。

未知語については、例えば、未知語辞書を作成し、MTP の辞書(2.3.3 参照)に登録する等の対応によって、正しく翻訳されることが期待できる。

そこで、特許庁では、令和元年度に、未知語に関する調査研究を実施した<sup>13</sup>。本調査研究を通じて、MTP の日英翻訳における約半年間の翻訳履歴から、翻訳失敗箇所を含む翻訳ログを抽出し、そこから数万語の未知語を検出して未知語辞書(日本語/英語)を作成した。なお、本事業を通じて令和2年3月に完成した未知語辞書は、既にMTP の辞書に追加登録済みである。

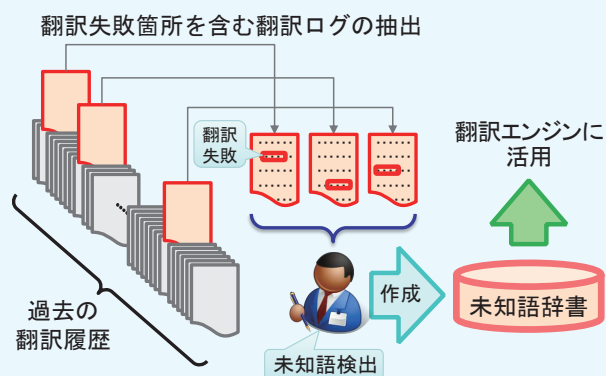


図10 未知語辞書の作成

## 4 おわりに

特許庁における特許情報の翻訳に係る取組みとして、これまで MTP を中心に述べてきたが、特許庁では他にも、例えば、中国審決の和訳文の提供<sup>14</sup>や、日本の審決等(審決、異議決定、判定)の英訳文の提供<sup>15</sup>等、様々な取組みを実施している。

また、海外に目を向けると、例えば WIPO (世界知的所有権機関)は、PCT の国際公開言語(日本語を含む10言語)の翻訳が可能な WIPO Translate<sup>16</sup> を利用した、自由入力文の機械翻訳サービスや PATENSCOPE における特許情報の機械翻訳サービス等を提供している。また、EPO (欧州特許庁)は、日本語を含む32言語の翻訳が可能な Patent Translate<sup>17</sup> を利用した、Espacenet における特許情報の機械翻訳サービス等を提供している。

WIPO Translate も Patent Translate も、MTP と同様、NMT がメインに利用されており、作成した機械翻訳文は一般利用者に無料で提供されている。ただし、照会の度に翻訳文が作成されるリアルタイム翻訳であるから、特許情報の翻訳文を対象としたテキスト検索はできない。

特許庁では、MTP の活用を含む特許情報の翻訳に関する種々の取組みを今後も実施する予定であるところ、前述した海外の動向もふまえ、より利用者のニーズに沿った翻訳に係るサービスを模索していきたい。

14 外国特許情報サービス「FOPISER (フォビサー)」収録内容追加のお知らせ(中国審決和訳文) <https://www.jpo.go.jp/support/fopiser/tsuika-20200622.html>

15 参考審決等の英訳 [https://www.jpo.go.jp/system/trial\\_appeal/info-shinketsu-eiyaku.html](https://www.jpo.go.jp/system/trial_appeal/info-shinketsu-eiyaku.html)

16 WIPO Translate <https://www.wipo.int/wipo-translate/>

17 Patent Translate <https://www.epo.org/searching-for-patents/helpful-resources/patent-translate.html>

13 新語対応による機械翻訳精度向上に関する調査事業