

機械翻訳技術の特許出願技術動向調査

Patent Trends in Machine Translation Technology

特許庁 総務部企画調査課知財動向班長

宮崎 大輔

平成 21 年 4 月特許庁入庁。審査第三部プラスチック工学、調整課審査企画室、審査第三部無機化学、特許情報室を経て、令和 3 年 1 月より現職。

1 特許出願技術動向調査について

特許出願は公報として広く一般に公開される。特許の公開情報は、企業・大学等における研究開発の成果に係る技術情報や権利情報である。これらの特許情報から、先端技術分野等の特許出願状況や研究開発の方向性を分析することで、企業、大学等における今後の研究開発の方向性を決定する指針となる。そこで、特許庁では、国の政策として推進すべき技術分野、社会的に注目されている技術分野等から技術テーマを選定して、平成 11 年度から「特許出願技術動向調査」と題し特許出願の分析を行っている。

令和 2 年度は、7 の技術テーマを選定し調査を行っており、本稿ではその中から、近年注目度の高い「機械翻訳」についての調査の概要を述べる。

2 機械翻訳

2.1 技術概要

「もう英語学習は要らなくなるかもしれない」という記事を見かけるようになった。その背景としては、ニューラルネットワーク技術を活用した、ニューラル機械翻訳と呼ばれる方式が 2014 年に登場し、その後、性能が格段に向上したことが挙げられる。

本調査では、この 2014 年を挟む期間における機械翻訳に関する特許出願技術動向を調査し、国内外の技術動向、日本及び外国の技術競争力の状況と今後の展望を明らかにすることを目的として、本技術に関する特許や

研究開発論文などの解析を行い、今後、取り組むべき課題や方向性について提言を行った。

機械翻訳の技術俯瞰図を図 1 に示す。本調査では、機械翻訳に関する技術を、その「翻訳対象」「データ取得・前処理」「翻訳方式・言語解析」「モデル構築・翻訳処理」「後処理・運用」「技術課題」の観点から分析した。

2.2 機械翻訳技術全体の特許出願動向

出願人国籍（地域）別ファミリー一件数比率¹及び推移を図 2 に示す。機械翻訳技術全体のファミリー一件数は、2016 年以降急激に増加している。国籍（地域）別では中国籍の増加が目立っている。

「出願人国籍（地域）」×「出願先国（地域）」の出願件数を図 3 に示す。米国は、いずれの国籍（地域）からも、最も多くの自国（地域）以外への出願がなされる出願先国となっており、国際競争の主戦場になっていると考えられる。米国籍は、いずれの出願先国（地域）へも、国（地域）外からの出願のうち最も多くの出願をしている国籍となっており、機械翻訳技術の牽引者になっていると考えられる。

1 「ファミリー一件数」とは、共通の優先権を持つ特許ファミリーを単位とした件数であり、いわゆる発明の数である。例えば、共通の優先権を持つ特許出願が日本と米国とにされた場合、「出願件数」は 2 件、「ファミリー一件数」は 1 件と数えられる。また、米国籍のファミリー一件数比率とは、国内外のファミリー一件数全体に占める米国籍によるものの比率を意味する。

機械翻訳に関する技術



図1 機械翻訳の技術俯瞰図

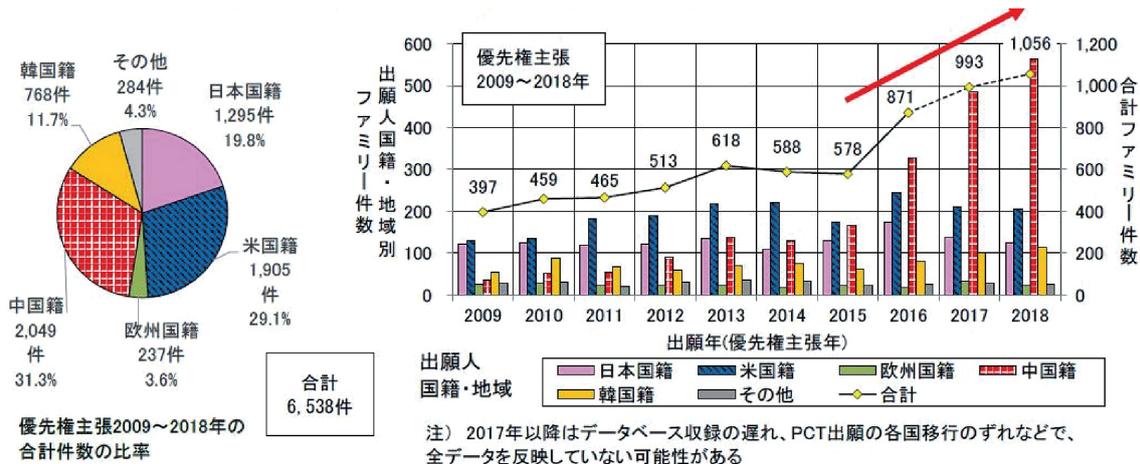


図2 出願人国籍(地域)別ファミリー件数比率及び推移(日米欧中韓 PCT への出願、優先権主張年: 2009-2018年)

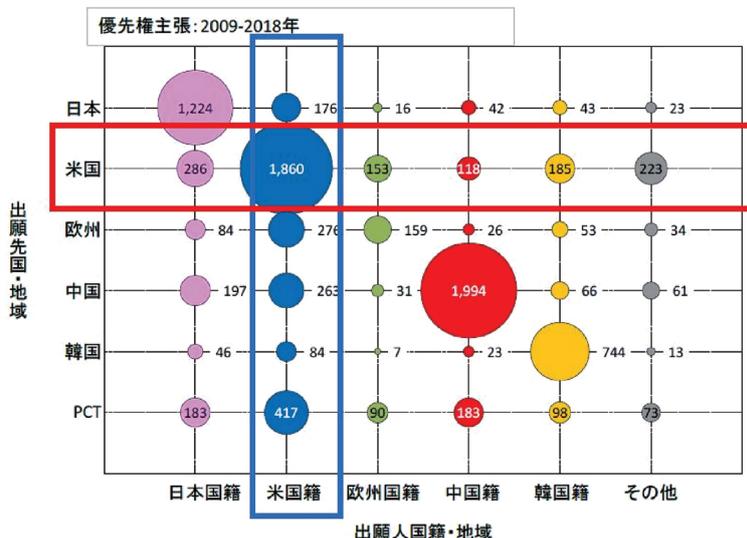


図3 (出願先: 日米欧中韓 PCT) 出願先国・地域別出願人国籍・地域別出願件数(出願年(優先権主張年): 2009-2018年)

2.3 ニューラル機械翻訳及び翻訳対象に関する出願動向

2.3.1 特許出願動向及びその分析

「ニューラル機械翻訳」に関する出願人国籍（地域）別ファミリー件数推移を図4に示す。「ニューラル機械翻訳」に関するファミリー件数は、2016年辺りを境に増加しているが、中国籍の件数増加が著しい一方、日本国籍の増加は遅れている。



図4 「ニューラル機械翻訳」出願人国籍（地域）別ファミリー件数推移（日米欧中韓 PCT への出願、優先権主張年：2009-2018年）

「翻訳対象」に関する技術区分における「米国籍のファミリー件数比率」×「日本国籍のファミリー件数比率」を図5に示す。各種「翻訳対象」に関する日本国籍のファミリー件数比率と米国籍のファミリー件数比率を比較すると、「特許・知財」は、日本国籍が特に優位な技術分野となっている。

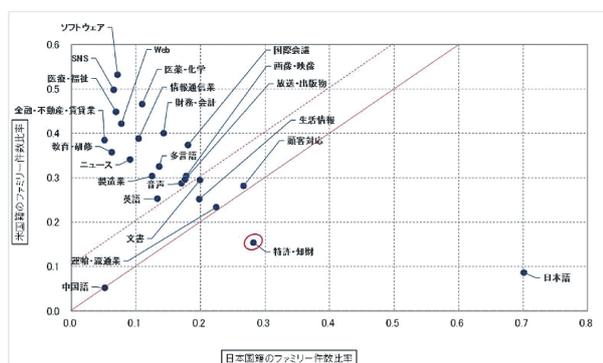


図5 「翻訳対象」に関する技術区分「米国籍のファミリー件数比率」×「日本国籍のファミリー件数比率」（日米欧中韓 PCT への出願、優先権主張年：2009-2018年）

デジタルデータの活用環境整備の取り組みの一例として、特許文献に関する取り組みを振り返ると、特許文献については、日本が世界に先駆けて、電子出願システム

を導入した（1990年～）²。

電子出願フォーマットについても、日本は2003年に特許・実用新案の出願様式のXML化を実現している。一方、米国特許庁は、2000年にXMLフォーマットによる電子出願制度を導入したが、国内で普及が進まないことから2006年にPDF形式をベースとした電子出願制度に変更した³。

XMLフォーマットは、少なくともPDF形式に比べれば、機械処理しやすいフォーマットであり、機械翻訳でもある程度利用しやすいが、PDF形式は人間にとっての読みやすさは高いものの、必ずしも機械翻訳で扱いやすいとは言えない。

そうした背景が、「特許・知財」分野における日本国籍の優位性につながっている可能性がある。

2.3.2 提言＜デジタルトランスフォーメーションの推進＞

日本企業は、「特許・知財」以外を対象とした分野においても同様の優位性を築き、デジタル経済社会における競争力を確保するため、機械翻訳において重要な基盤となるデジタルデータの活用環境の整備（機械処理しやすいデジタルデータの生成・共有）を図っていくべきである。

2.4 技術課題及び後処理・運用に関する出願動向、翻訳対象と付与率の関係

2.4.1 特許出願動向及びその分析

「技術課題」に関する技術区分における「米国籍のファミリー件数比率」×「日本国籍のファミリー件数比率」を図6に示す。「分野適応」（分野特有の用語や表現、背景や文脈に応じた訳出ができるようにすること）や、情報セキュリティの担保に関する「秘匿性」を「技術課題」とした特許出願を見ると、機械翻訳技術を牽引している米国籍が特に優位な技術分野と位置づけられる。

「後処理・運用」に関する技術区分における「米国籍のファミリー件数比率」×「日本国籍のファミリー件数比率」を図7に示す。「適応型機械翻訳」（顧客向けにカスタマイズした機械翻訳）に関する特許出願を見ると、

2 特許庁（2017）「特許庁の情報システムの概要」

3 特許庁「グローバルな知財IT環境の構築」Japio 2009 YEAR BOOK

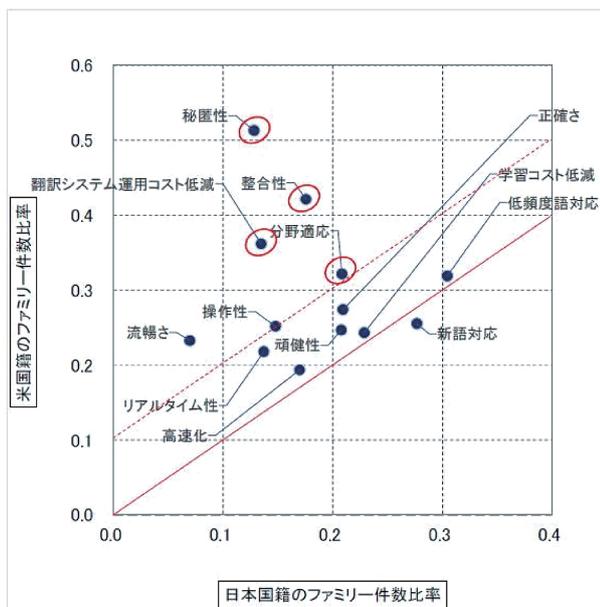


図6 「技術課題」に関する技術区分「米国籍のファミリー件数比率」×「日本国籍のファミリー件数比率」(日米欧中韓 PCT への出願、優先権主張年：2009-2018年)

米国籍が日本国籍より特に優位な技術分野となっている。

訳に対するニーズが高いと予測されている翻訳対象である「情報通信業」(エレクトロニクスに対応する技術区分と考えられる)、「ソフトウェア」、「Web」は、米国籍の付与率が日本国籍より高くなっている。

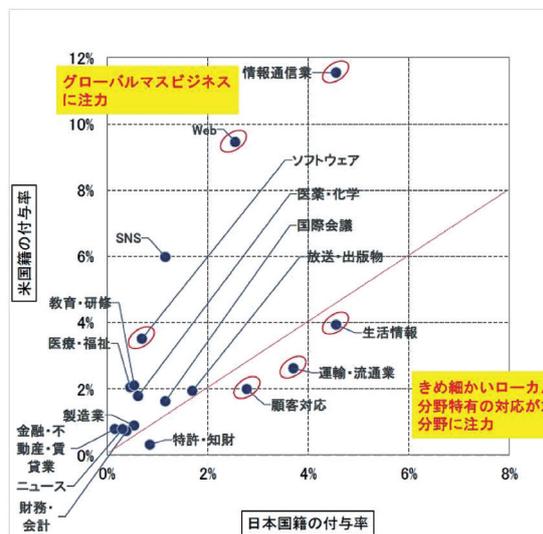


図8 「翻訳対象」に関する技術区分「米国籍の付与率」×「日本国籍の付与率」(日米欧中韓 PCT への出願、優先権主張年：2009-2018年)

注) 「農林水産業」「鉱業・建設業」「製品説明」「電気・ガス・水道業」「法務」は国内外全体のファミリー件数が少ないためプロットから除外した。

一方、「運輸・流通業」(運輸、eコマースに対応する技術区分と考えられる)、「顧客対応」(顧客とのコミュニケーションに対応する技術区分と考えられる)、「生活情報」(ホスピタリティーに対応する技術区分と考えられる)の付与率は日本国籍の方が高くなっている。

2.4.2 提言<機械翻訳を活用した翻訳のビジネスサービスとしての付加価値向上>

日本企業は、機械翻訳を活用した翻訳のビジネスサービスとしての付加価値向上のため、「分野適応」や「適応型機械翻訳」のサービス化に注力すべきである。ビジネスユースの機械翻訳サービスにおいては、情報セキュリティの担保も重要となる。

対象分野としては、「運輸・流通業」「顧客対応」「生活情報」等、機械翻訳市場におけるニーズが高く、かつ、大部分が自動処理によるグローバルマスビジネスよりも、人手による対応も含めたきめ細かいローカルな対応、分野特有の対応が求められる分野は、これまで日本企業が注力してきており、蓄積がある分野であり、引き続き

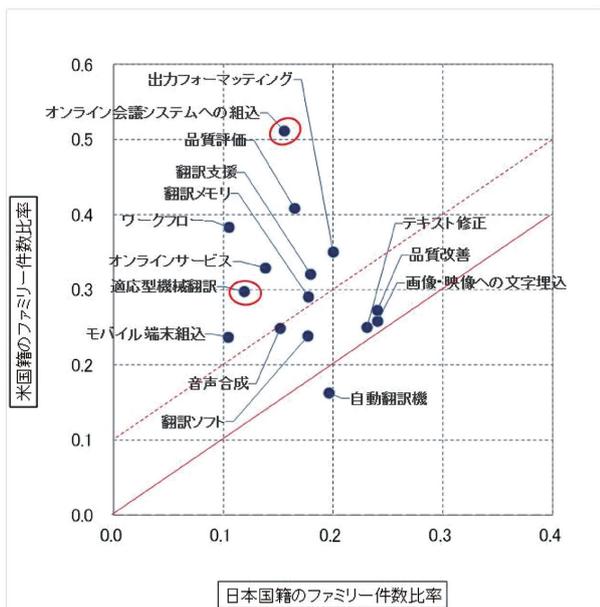


図7 「後処理・運用」に関する技術区分「米国籍のファミリー件数比率」×「日本国籍のファミリー件数比率」(日米欧中韓 PCT への出願、優先権主張年：2009-2018年)
注) 「前処理の学習」は国内外全体のファミリー件数が少ないためプロットから除外した。

「翻訳対象」に関する技術区分における「米国籍の付与率⁴⁾」×「日本国籍の付与率」を図8に示す。機械翻

4) ここで、技術区分Aに関する米国籍の付与率とは、米国籍の全体のファミリー件数の中で技術区分Aが付与されているものの割合を意味する。

注力していくべきである。

2.5 同時翻訳の実現に関する出願動向

2.5.1 特許出願動向及びその分析

同時通訳や文脈・話者の意図等の加味に関連する技術区分におけるファミリー件数推移を図9に示す。機械翻訳の同時通訳への適用に関する特許出願を見ると、「音声」を入力とする機械翻訳、その「リアルタイム処理」「音声認識」に関する特許出願は、ファミリー件数が2016年辺りから急激に増加していることから、成長分野と考えられ、今後の一層の推進が期待される。

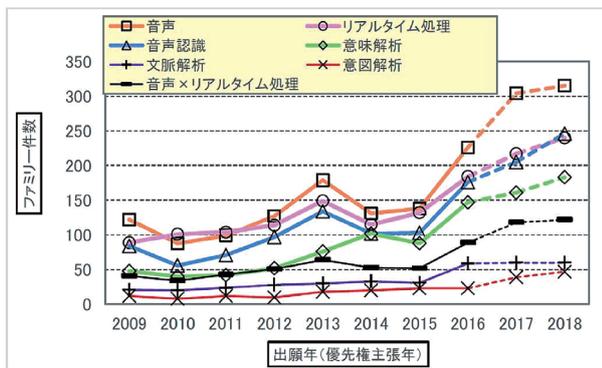


図9 同時通訳や文脈・話者の意図等の加味に関連する技術区分—ファミリー件数推移（日米欧中韓 PCT への出願、優先権主張年：2009-2018年）

音声翻訳の高度化に関連する技術区分におけるファミリー件数推移を図10に示す。「ノイズ除去」「声の特徴認識」「同時通訳用コーパス活用」「同時通訳用コーパス活用」「同時通訳用コーパス活用」も近年増加の兆しが見られる。

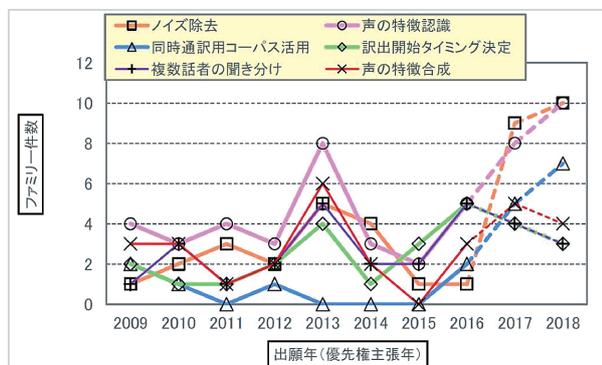


図10 音声翻訳の高度化に関連する技術区分—ファミリー件数推移（日米欧中韓 PCT への出願、優先権主張年：2009-2018年）

「翻訳方式・言語解析」に関する技術区分における「米国籍のファミリー件数比率」×「日本国籍のファミリー

件数比率」を図11に示す。文脈・話者の意図等を補う同時通訳（議論レベル）の実現に関連する、「意味解析」「文脈解析」「意図解析」に関する特許は、米国籍が特に優位な技術分野と位置づけられる。

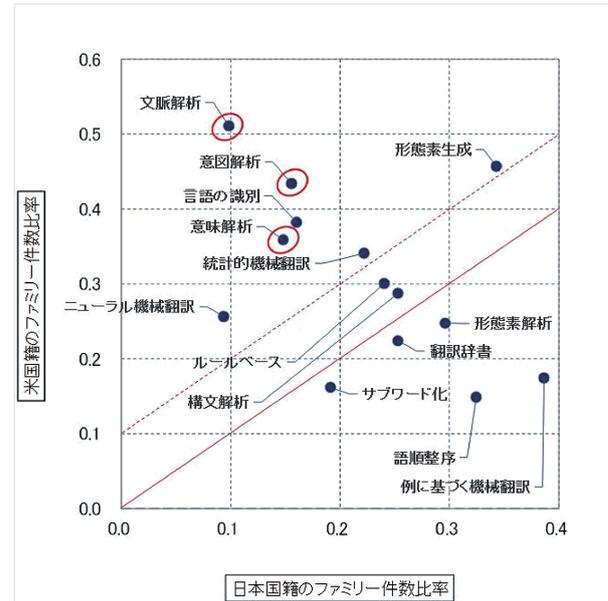


図11 「翻訳方式・言語解析」に関する技術区分—「米国籍のファミリー件数比率」×「日本国籍のファミリー件数比率」（日米欧中韓 PCT への出願、優先権主張年：2009-2018年）

「データ取得・前処理」に関する技術区分における「米国籍のファミリー件数比率」×「日本国籍のファミリー件数比率」を図12に示す。「マルチモーダル」に関する

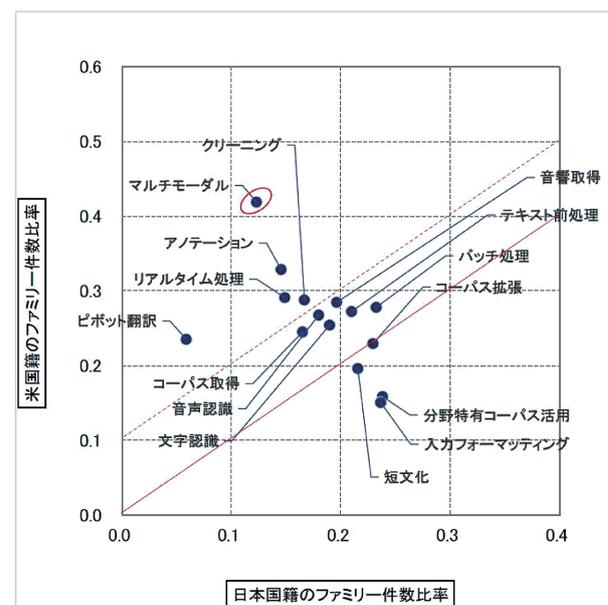


図12 「データ取得・前処理」に関する技術区分—「米国籍のファミリー件数比率」×「日本国籍のファミリー件数比率」（日米欧中韓 PCT への出願、優先権主張年：2009-2018年）

る特許出願を見ると、米国籍が特に優位であり、日本国籍の技術蓄積が少ない。

また、図7において、機械翻訳技術の「オンライン会議システムへの組込」に関する特許出願を見ると、米国籍が特に優位であり、日本国籍の技術蓄積が少ない。

2.5.2 提言<言語の壁がない生活・ビジネス環境構築に向けた同時通訳の技術開発>

日本企業は、言語の壁がない生活・就労・ビジネスコミュニケーション環境を構築するため、「音声」を入力とする機械翻訳やその「リアルタイム処理」など、機械翻訳の同時通訳への適用を一層推進していくべきである。

そのため、チャレンジングな課題であるものの、「意味解析」「文脈解析」「意図解析」を踏まえた、文脈・話者の意図等を補う同時通訳（議論レベル）の実現や、「音声認識」「同時通訳用コーパス活用」「ノイズ除去」「声の特徴認識」「マルチモーダル」などの技術開発を強化していくべきである。

また、感染症拡大対策のためにも、機械翻訳技術の「オンライン会議システムへの組込」を進めるべきである。

3 結び

本稿では、令和2年度に調査を実施したテーマの中から、「機械翻訳」の調査結果を紹介した。なお、ページ数の都合上本稿では触れなかったが、本調査では上記の提言のほかに、機械翻訳を活用した翻訳業務の生産性向上についても提言を行っている。

特許出願動向調査の要約版は特許庁ウェブサイトに掲載されており、報告書については、国立国会図書館、特許庁図書館で閲覧可能である。

(<https://www.jpo.go.jp/resources/report/gidou-houkoku/tokkyo/index.html>)

我が国の企業、大学等が、研究開発戦略策定の際、特許出願技術動向調査の調査結果を有効に活用することで、効率的な技術開発を進め、結果として我が国の国際競争力強化につながれば幸甚である。