

令和5年度
特許情報の機械翻訳における多言語対応に
向けた課題検討に関する調査事業
〈調査報告書〉

令和6年3月

TOPPAN 株式会社

本報告書は、特許庁委託調査として、TOPPAN株式会社が実施した「令和5年度特許情報の機械翻訳における多言語対応に向けた課題検討に関する調査事業」の成果を取りまとめたものです。

目次

| | | |
|--------|---------------------------------------|----|
| 1. | はじめに | 1 |
| 1.1. | 調査の目的・背景..... | 1 |
| 1.2. | 調査の内容 | 1 |
| 1.2.1. | 調査全体の構成..... | 1 |
| 1.2.2. | 調査対象言語および調査対象国について..... | 2 |
| 1.2.3. | 各調査項目の概要..... | 3 |
| 2. | 機械翻訳サービス・エンジンの種類・内容に関する調査..... | 5 |
| 2.1. | 調査の概要 | 5 |
| 2.2. | 国内外で提供されている汎用の機械翻訳サービス及びエンジンについて..... | 5 |
| 2.2.1. | 汎用機械翻訳エンジン調査の全体概略..... | 5 |
| 2.2.2. | 事例1：Google Translate | 7 |
| 2.2.3. | 事例2：Microsoft Translator..... | 10 |
| 2.2.4. | 事例3：DeepL..... | 12 |
| 2.2.5. | 事例4：SAP Translation Hub..... | 14 |
| 2.2.6. | 事例5：SYSTRAN（シストラン） | 16 |
| 2.2.7. | 事例6：みらい翻訳..... | 18 |
| 2.2.8. | 事例7：みんなの自動翻訳..... | 20 |
| 2.2.9. | 事例8：T-400 | 22 |
| 2.3. | 特許情報に特化した機械翻訳エンジン | 24 |
| 2.3.1. | 各国知財庁の機械翻訳の提供状況..... | 24 |
| 2.3.2. | 特許特化型機械翻訳サービス・エンジンの事例..... | 30 |
| 3. | 日英中韓以外の言語をネイティブとする知財庁の状況に関する調査..... | 38 |
| 3.1. | 調査の概要 | 38 |
| 3.1.1. | 調査の目的..... | 38 |
| 3.1.2. | 調査項目等..... | 38 |
| 3.2. | アンケート結果分析..... | 38 |
| 3.2.1. | 特許情報サービスにおける日英翻訳機能について..... | 38 |
| 3.2.2. | 公報・審査結果の英語での理解..... | 43 |
| 3.2.3. | 審査業務と英語力..... | 45 |
| 3.3. | アンケート結果からの示唆..... | 46 |
| 4. | 調査対象言語を使用する国々に関する経済的情報に関する調査..... | 47 |
| 4.1. | 調査の概要 | 47 |

| | | |
|---------|---------------------------------------|-----|
| 4.2. | 基礎的経済データについて..... | 50 |
| 4.2.1. | 調査対象国の経済規模（GDP、成長率、輸出入額等）..... | 50 |
| 4.2.2. | 海外直接投資および技術貿易収支..... | 58 |
| 4.2.3. | 特許関連のデータ..... | 61 |
| 4.3. | その他の経済的データについて..... | 65 |
| 4.3.1. | スペイン..... | 65 |
| 4.3.2. | メキシコ..... | 67 |
| 4.3.3. | ポルトガル..... | 70 |
| 4.3.4. | ブラジル..... | 73 |
| 4.3.5. | イタリア..... | 76 |
| 4.3.6. | インドネシア..... | 79 |
| 4.3.7. | マレーシア..... | 82 |
| 4.3.8. | ベトナム..... | 85 |
| 4.3.9. | タイ..... | 88 |
| 4.3.10. | サウジアラビア..... | 91 |
| 4.3.11. | インド..... | 95 |
| 4.3.12. | トルコ..... | 98 |
| 4.4. | 調査対象地域（ASEAN、中南米、中東等）の総括..... | 101 |
| 4.4.1. | 海外における知的財産活用支援の枠組み..... | 101 |
| 4.4.2. | Country Factsheets における特許に係る言語情報..... | 105 |
| 4.4.3. | ASEAN、中南米、中東における民間のニーズ..... | 106 |
| 5. | 調査対象言語に関するコーパス作成の技術的情報..... | 109 |
| 5.1. | 調査の概要..... | 109 |
| 5.2. | パテントファミリーデータの取得元およびデータの状態..... | 109 |
| 5.2.1. | スペイン語..... | 109 |
| 5.2.2. | ポルトガル語..... | 111 |
| 5.2.3. | イタリア語..... | 113 |
| 5.2.4. | インドネシア語..... | 114 |
| 5.2.5. | マレー語..... | 116 |
| 5.2.6. | ベトナム語..... | 116 |
| 5.2.7. | タイ語..... | 118 |
| 5.2.8. | アラビア語..... | 119 |
| 5.2.9. | ヒンディー語..... | 121 |
| 5.2.10. | トルコ語..... | 122 |
| 5.3. | 多言語 OCR ツール..... | 123 |
| 5.3.1. | 公開情報調査・ヒアリング調査..... | 123 |
| 5.3.2. | 多言語 OCR 調査項目..... | 123 |
| 5.3.3. | 公開情報調査・ヒアリング調査結果..... | 124 |

| | | |
|---------|--------------------------------|-----|
| 5.3.4. | 公開情報調査・ヒアリング調査考察..... | 125 |
| 5.3.5. | 試作コーパス言語の OCR 精度検証..... | 126 |
| 5.3.6. | 精度検証設定..... | 126 |
| 5.3.7. | 精度検証結果..... | 131 |
| 5.3.8. | 精度検証結果考察..... | 132 |
| 5.4. | 文分割ツール..... | 133 |
| 5.4.1. | スペイン語..... | 133 |
| 5.4.2. | ポルトガル語..... | 133 |
| 5.4.3. | イタリア語..... | 134 |
| 5.4.4. | インドネシア語..... | 134 |
| 5.4.5. | マレー語..... | 134 |
| 5.4.6. | ベトナム語..... | 134 |
| 5.4.7. | タイ語 | 135 |
| 5.4.8. | アラビア語..... | 135 |
| 5.4.9. | ヒンディー語..... | 135 |
| 5.4.10. | トルコ語..... | 136 |
| 5.4.11. | 多言語対応ツール..... | 136 |
| 5.5. | 形態素解析（単語分割）ツール..... | 137 |
| 5.5.1. | スペイン語..... | 137 |
| 5.5.2. | ポルトガル語..... | 137 |
| 5.5.3. | イタリア語..... | 137 |
| 5.5.4. | インドネシア語..... | 137 |
| 5.5.5. | マレー語..... | 137 |
| 5.5.6. | ベトナム語..... | 138 |
| 5.5.7. | タイ語 | 139 |
| 5.5.8. | アラビア語..... | 139 |
| 5.5.9. | ヒンディー語..... | 139 |
| 5.5.10. | トルコ語..... | 140 |
| 5.5.11. | 多言語対応ツール..... | 140 |
| 5.6. | アライメントツール..... | 141 |
| 5.7. | 対訳辞書作成ツール..... | 143 |
| 5.8. | パテントファミリーデータからの対訳コーパス作成手順..... | 144 |
| 5.9. | コーパス試作、精度評価、作成可能な対訳コーパス量..... | 146 |
| 5.9.1. | 試作コーパスのアライメント精度評価の概要..... | 146 |
| 5.9.2. | アライメント精度評価の実施方法..... | 147 |
| 5.9.3. | インドネシア語..... | 150 |
| 5.9.4. | ベトナム語..... | 155 |
| 5.9.5. | タイ語 | 160 |

| | |
|---|-----|
| 5.10. その他の関連技術..... | 165 |
| 5.10.1. 多言語で事前訓練された大規模なモデルを用いた多言語処理について | 165 |
| 5.10.2. 対訳コーパスのデータ拡張..... | 166 |
| 6. 調査結果の総合分析..... | 168 |

1. はじめに

1.1. 調査の目的・背景

日米欧中韓の知財庁（以下「五庁」）への特許出願は、世界の特許出願件数の大半を占めており（2021年は85.1%）、日本企業等の海外展開を支援する際、五庁の特許情報を提供することは不可欠である。特許庁ではこれまでに、対訳コーパスを学習した機械翻訳エンジンを用い、日本の審査情報等の海外発信において日英機械翻訳、外国公報の検索照会において中日・韓日機械翻訳を提供してきた。

一方で、世界における日本企業の現地法人数について、ASEANの割合が2020年まで10年連続で拡大している。日本からASEAN各国への特許出願件数が増加傾向を示すとともに、中南米、インド、中東等への特許出願件数についても一定の出願が毎年見られる。このように、日本企業の海外ビジネス展開にあたって、米欧中韓以外の国への関心が高まり、これが維持されている。

五庁で主に使用される言語（日英中韓）以外の言語を対象とした機械翻訳の精度を高め、五庁以外の知財庁の特許情報の日本語への機械翻訳、日本の審査情報等の現地語への機械翻訳を提供することが望まれるが、日英中韓以外の言語について、機械翻訳の精度に資する対訳コーパス作成は国内外で進んでいない。

これらから、本調査では、国内外で提供される機械翻訳エンジンの現状、対訳コーパス作成の対象となり得る候補言語を使用する国の経済的情報、候補言語の対訳コーパス作成における技術的情報等を収集し、特許情報の機械翻訳の訳質向上に取り組む際の基礎データ・資料を作成した。

1.2. 調査の内容

1.2.1. 調査全体の構成

本調査では、①機械翻訳エンジンの種類・内容、②日英中韓以外の言語をネイティブとする知財庁の状況、③調査対象言語を使用する国々に関する経済的状況、④調査対象言語に関するコーパス作成の技術的情報の4つの項目に対する調査を実施した。

各調査項目間の関係は、図1のとおりである。調査の実施にあたっては、機械翻訳エンジンの言語対応状況、日本から各国への出願件数、日本企業の現地法人数等の情報から、調査対象言語と調査対象国を選定した後に、調査項目1～4を開始した。調査項目4の対訳コーパスの試験的作成にあたっては、調査項目1～4で得られた調査結果を総合的に勘案し、対象言語を選定した。

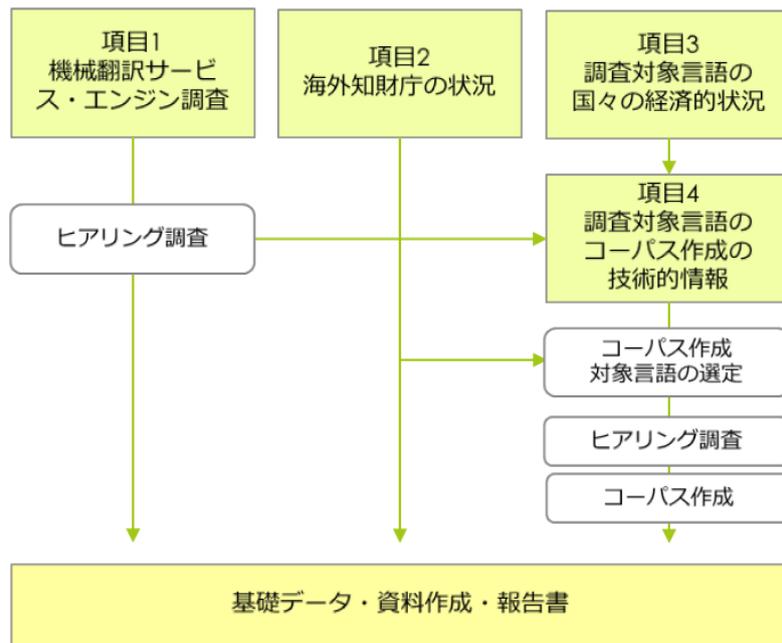


図1 調査の全体構成

1.2.2. 調査対象言語および調査対象国について

本調査の対象言語および対象国の選定にあたっては、特許情報に特化した機械翻訳エンジンの調査において各国知財庁の対象先（調査項目1）、外国特許情報サービス（FOPISER）における特許・実用新案の検索・照会等の実施国・希望国¹、日本との経済的関係及び知財等の注目国（調査項目3）の3つの観点から選定を行った。

本調査の対象言語の選定においては、翻訳元言語又は翻訳先言語に英語又は日本語を含むことを前提とした上で、スペイン語、ポルトガル語、イタリア語、インドネシア語、マレー語、ベトナム語、タイ語、アラビア語、ヒンディー語、トルコ語の10言語を対象とした。また、調査対象国については、調査対象言語となる12カ国（スペイン、メキシコ、ポルトガル、ブラジル、イタリア、インドネシア、マレーシア、ベトナム、タイ、サウジアラビア、インド、トルコ）を選定した。

¹ 特許庁（2022）「令和3年度 特許情報提供サービスの現状と今後に関する調査」。
 〈<https://www.jpo.go.jp/resources/report/sonota/service/document/index/2021hokokusho.pdf>〉

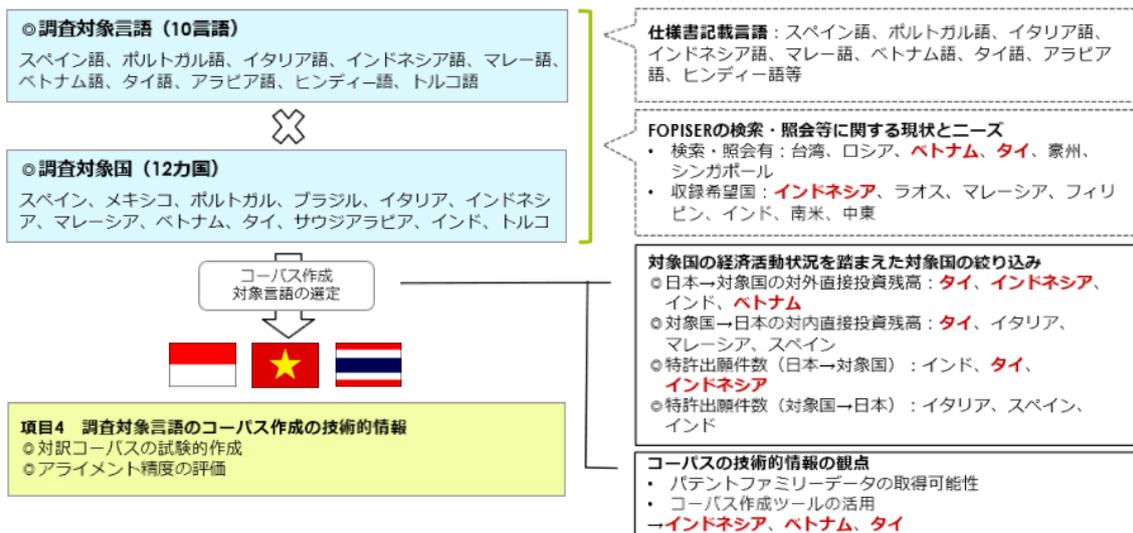


図2 コーパス作成言語の選定プロセス

1.2.3. 各調査項目の概要

(1) 項目1：機械翻訳サービス・エンジンの種類・内容

本項目は、①国内外で提供されている汎用の機械翻訳エンジン、②特許情報に特化した機械翻訳エンジンからなる。

国内外で提供されている汎用の機械翻訳エンジンでは、民間事業者、又はアカデミアが提供する機械翻訳エンジンについて、使用アルゴリズム、対象言語、学習コーパスの種類、ユーザーによる追加学習機能、訳質、提供開始時期、料金等についての内容を調査した。

特許情報に特化した機械翻訳エンジンでは、各国知財庁（JPO、USPTO、EPO、CNIPA、KIPO、UKIPO、WIPO、EUIPO等）のほか、特許特化型の機械翻訳サービスを展開する事業者を対象に、特許情報に特化した機械翻訳サービス・エンジンについて、サービス概要、機械翻訳エンジン・アルゴリズム、対象言語、学習コーパスの種類、ユーザーによる追加学習、訳質、提供開始時期、料金の内容を調査した。

(2) 項目2：日中韓以外の言語をネイティブとする特許庁の状況

特許審査ハイウェイ（PPH）に限らず、各国特許庁審査官は、特許審査を行うにあたって、同等の内容で他国に出願され、先に他国の審査官によって特許審査が行われた結果を参照し、特許庁間でのワークシェアリングを行っている。

他国の公報・審査結果を参照する場合、英語以外の言語は、基本的には英語に翻訳されてその内容を理解することが多いと考えられ、英語以外の言語をネイティブとする審査官は、既に業務において英語の読み書きに精通していることが考えられる。

調査では、海外の知財庁の審査官を対象に、ネイティブ言語で公報・審査結果を参照する場合と比較し、英語で他国の公報・審査結果を参照する場合の理解度、問題点（訳質、英語力等）、日本国特許庁が提供する日英機械翻訳サービスへの評価、他庁への要望等について、アンケート調査（電子メール）を実施した。調査対象は、日英中韓以外の言語をネイティブとする海外の知財庁の審査官であり、調査の実施にあたっては、日本国特許庁を介して、審査官に調査協力の依頼を行った。最終的に 10 カ国からの回答を得られた。

(3) 項目 3：調査対象言語を使用する国々に関する経済的状況

調査対象国を対象に、基礎的経済データ（経済規模、日本との間の対外投資・技術貿易収支等）、日本企業の進出状況、知財に係る日本からの出願件数（PPH 申請件数を含む）やその他の経済的情報を収集した。調査にあたっては、定量的な経済的状況の情報以外に、日本企業のビジネス展開、又は日本国の外交における地理的・地政学的な状況に関する定性的な情報も収集した。

(4) 項目 4：調査対象言語に関するコーパス作成の技術的情報

項目 1 で対象とした言語について、特に特許情報に係るコーパス作成の技術的情報として、パテントファミリーを利用したコーパス作成の具体的な候補手順（OCR・文分割・形態素解析・文アライメントに用いるツール等）とそれらの課題（データ取得の困難性を含む）、候補手順を採用したときの推定コストやその他の事項を、（例えば、コーパスの試験的な作成等を通して）調査する。その際には、「平成 28 年度ベトナム・タイ語の対訳コーパス・辞書の自動作成に向けたツール等の検証調査（特許庁 HP で公開）」及び「自然言語処理（一般社団法人言語処理学会が発行）、特許機械翻訳の課題解決に向けた機械翻訳技術解説、2022 年 29 巻 3 号、p.925 985」を参考に調査を行った。

2. 機械翻訳サービス・エンジンの種類・内容に関する調査

2.1. 調査の概要

本項目では、1.2.3. (1) で示したとおり、民間事業者、又はアカデミアが提供する汎用機械翻訳エンジンについて、使用アルゴリズム、対象言語（何言語から何言語という翻訳の方向性を含む）、学習コーパスの種類（いかなる文書からどのように作成されたものか）、ユーザーによる追加学習が可能か、訳質、提供開始時期、料金、その他の観点から、公開情報調査を実施した（以下、汎用機械翻訳エンジン調査）。

また、特許情報に特化した機械翻訳エンジンを対象に、日本国特許庁（JPO）、米国特許商標庁（USPTO）、欧州特許庁（EPO）、中国国家知識産権局（CNIPA）、韓国特許庁（KIPO）、英国知的財産庁（UKIPO）、世界知的所有権機関（WIPO）、欧州連合知的財産庁（EUIPO）の8つの知財庁における機械翻訳サービスの提供状況を調査するとともに、民間事業者が提供する機械翻訳エンジンについて、汎用機械翻訳エンジンと同様に使用アルゴリズム、対象言語、学習コーパスの種類、追加学習機能、訳質、提供開始時期、料金、その他の観点から、公開情報調査を実施した（以下、特許特化型機械翻訳エンジン調査）。

機械翻訳エンジンは、機械翻訳サービスの提供形態により、様々なエンジンを組み合わせて提供されている。「機械翻訳エンジン〈単体〉－機械翻訳サービス」ではなく、「機械翻訳サービス－機械翻訳エンジン〈複数〉」で提供されているため、本項目の調査にあたっては、機械翻訳サービスの情報を収集・整理し、それらサービスで搭載される機械翻訳エンジンの情報を調査した。

調査の実施にあたっては、情報収集項目のうち、使用アルゴリズム、学習コーパスの種類、訳質等の情報は、機械翻訳エンジンの提供事業者が公表している情報には限りがあるため、情報収集にあたっては、二次情報（記事等）を含めて情報収集を行った。また、各機械翻訳サービスが提供する言語の情報については、本調査対象言語（10言語）を対象とした。

2.2. 国内外で提供されている汎用の機械翻訳サービス及びエンジンについて

2.2.1. 汎用機械翻訳エンジン調査の全体概略

汎用機械翻訳エンジン調査では、Google Translate、Microsoft Translator、DeepL、SAP Translation Hub、SYSTRAN、みらい翻訳、みんなの自動翻訳、T-400の8つの機械翻訳サービスを対象に実施した。

表1 調査対象の汎用機械翻訳サービス一覧、

| サービス名 | 概略 |
|------------------|---|
| Google Translate | 2015年からニューラル機械翻訳を開始。Google CloudのTranslation AI |

| サービス名 | 概略 |
|--|--|
| | では、高速で動的な機械翻訳機能を使用し、コンテンツやアプリの多言語化の対応を行っている。提供機能は、Cloud Translation、Translation Hub、Translation API (Basic、Advanced) 、AutoML Translation、Media Translation である。 |
| Microsoft Translator | Microsoft Cognitive Service の一部。企業向けクラウドサービスも提供している。翻訳製品は、テキスト翻訳 (2007 年から使用され、2011 年からは顧客向けの API を提供)、ドキュメント翻訳、翻訳カスタマイズ (Custom Translator を利用することで、企業、アプリ開発者、言語サービスプロバイダは、自社のビジネスや業界で使用されている専門用語を理解したニューラル翻訳システムを構築することができる) を提供している。Translator は現在、ニューラルネットワーク (LSTM) の翻訳品質向上の新しい翻訳を提供している。 |
| DeepL | 2016 年に、Jaroslaw Kutylowski 博士が創業し、2017 年 8 月にドイツ・ケルンで DeepL SE 事業を展開。31 言語をサポートしている。DeepL Proha、ビジネス向けのサービスであり、自然的な訳文を提供。世界中で 20000 社以上が活用し、DeepL Pro では、法務チームで使用可能なセキュアな AI 翻訳機能を提供している。EU の公文書を使って学習した AI が、特に法律関連のテキストで高精度な訳文を出力している。 |
| SAP Translation Hub | SAP が自社製品のユーザー向けに開発した、機械翻訳クラウドプラットフォーム。SAP Translation Hub では、SAP が承認した翻訳と用語のリポジトリ、および機械翻訳を使用して、UI テキストおよび文書を翻訳する。翻訳のリソースには、広範囲の API メソッドを利用して、または統合されたワークフローシナリオを UI で使用して、アクセスできる。さらに、開発者は、開発フェーズで翻訳元テキストに関する提案を得ることが可能である。 |
| SYSTRAN | 1968 年に創業。自動翻訳に特化した研究開発企業であり、OECD、米国連邦政府インテリジェンス・コミュニティ、仏・国防省、NATO/OTAN 等の機関でも利用されている。翻訳エンジンは、PNMT™ (Pure Neural™ Machine Translation)。カスタマイズ可能で、法律、自動車産業、IT、観光等に至るまで、各顧客の特定のコンテキストに適応した機械翻訳を提供している (専門分野のエンジン: 製造、金融、医療・製薬、法務、IT、テクノロジー、教育、農業、特許、ビジネス、対話、汎用エンジン)。 |
| Mirai Translator (みらい翻訳) *株式会社未来翻訳 | 企業内の文書翻訳の AI 自動翻訳ソリューションとして提供している。翻訳エンジンは、NICT の研究成果を活用したものである。分野専門モデルの搭載 (契約書・規定等の法務関連の文書、決算短信・有価証券報告書等の財務関連文書の翻訳に適した法務・財務モデル、特許公報・明細書等の特許書類特有の言い回し・長文に対応した特許モデルが利用可能)。 |
| みんなの自動翻訳 | 情報通信研究機構 (NICT) が開発した自動翻訳技術。AI 技術で多用される深い階層構造を持つネットワークを用いた自動翻訳技術 (AI 自動翻訳) である。「みんなの自動翻訳」は、TexTra として、Word や PowerPoint のファイルを直接翻訳する機能を搭載している。 |
| T-400 *ロゼッタ社 | 産業機械・電子部品・製薬分野などのガイドライン・公的文書を大量に学習させた翻訳エンジンを有する自動翻訳サービスである。2,000 分野から選択/検索でき、各分野には国内外の公的文書・ガイドライン等が収録されている。プロの翻訳者の意見を取り入れて編集機能を搭載している。 |

各機械翻訳サービスの調査項目は、(1) サービスの概要、(2) 翻訳エンジン・アルゴリズム・学習コーパス・訳質、(3) 対象言語、(4) 追加機能等、(5) その他からなる。

サービスの概要では、サービスの提供形態、特徴等を取りまとめた。

翻訳エンジン・アルゴリズム・学習コーパス・訳質の項目は、そのまま、翻訳エンジン、アルゴリズム、学習コーパス、訳質の 4 つの中項目に分割した。汎用機械翻訳サービスの提供事業者が公開する情報は、翻訳エンジンではエンジン名、開発者、サービスへの提供形態等にとどまる他、アルゴリズム及び学習コーパスでは、詳細情報はほとんど公開されていない²。訳質については、一部のサービスで自社評価値を公表している。

対象言語については、本調査で対象とする 10 言語の対応状況を把握した。

追加機能等については、追加学習機能についての情報を整理した。追加機能の多くは、分野別の学習オプションと、企業内データベース（社内表現、言い回し等）の学習機能等が一般的である。

その他の項目については、提供開始時期、料金、その他の情報を収集・整理した。

表 2 汎用機械翻訳サービス及びエンジンの調査項目及び情報の公開性

| 調査項目 | | 公開情報の有無 |
|-------------------------|--------|------------------------------|
| サービスの概要 | — | 公開 |
| 翻訳エンジン・アルゴリズム・学習コーパス・訳質 | 翻訳エンジン | 非公開（エンジン名まで） |
| | アルゴリズム | 非公開（機械翻訳形式：Transformer 型等まで） |
| | 学習コーパス | 一部公開（データ分野、辞書機能まで） |
| | 訳質 | 機械翻訳サービスの提供事業者による自己評価等 |
| 対象言語 | 提供言語 | 公開 |
| | 調査対象言語 | 公開 |
| 追加機能等 | 追加学習 | 一部公開 |
| その他 | 提供開始時期 | 公開 |
| | 料金 | 公開 |
| | その他 | — |

2.2.2. 事例 1 : Google Translate³

(1) サービスの概要

Google Translate は、2015 年からニューラル機械翻訳サービスの提供を開始した。Google

² アルゴリズムに関する論文情報は明示されていることもあるが、一方で、提供サービスに含まれる機械翻訳エンジンにどのように組み込まれているかは定かでないため、公開情報上、アルゴリズムが用いられているかどうかはブラックボックスである。学習コーパスについては、特許情報に係る特定のデータセットを組み込んでいる等の情報はあがるが、量的、質的な投入量に関する情報は公開されていないため、学習コーパスに関する情報は、対象分野に留まる。

³ Google Translate 〈<https://translate.google.com/about/?hl=ja>〉

Cloud の Translation AI (Cloud Translation) では、高速で動的な機械翻訳機能を使用し、コンテンツやアプリの多言語化の対応を行っている。

提供機能 (サービス) は、「Cloud Translation」、「Translation Hub」、「Translation API (Basic、Advanced)」、「AutoML Translation」、「Media Translation」である。

「Cloud Translation」は、ウェブサイトやアプリケーションで、API を使用してプログラムによってテキストを動的に翻訳するもので、デフォルトの翻訳システムは、Google の事前トレーニング済みのニューラル機械翻訳 (NMT) モデルを使用している。当該機能は、ベーシック、アドバンスの 2 つが提供され、アドバンスには後述する AutoML カスタムモデルや、バッチ リクエスト、用語集などの機能が含まれている。アドバンスは、医療分野やマーケティング コンテンツなど、特定のスタイルに従う必要のあるコンテンツを定期的に翻訳する場合に用いることが想定され、「AutoML Translation」でモデルのトレーニングをすることができる⁴。

「Translation Hub」では、大量のドキュメントをさまざまな言語に翻訳する組織向けのセルフサービス ドキュメント翻訳サービスである。具体的な利用形態の例として、行政機関が公開している公衆衛生に関する情報を本機能により、多言語に翻訳できるほか、翻訳メモリを使用することで、ドキュメントの翻訳とレビュー (人間によるフィードバック) を進めながら翻訳結果を今後の翻訳に再利用することができる。

「AutoML Translation」は、分野の専門知識を持つ開発者、翻訳者、ローカライズ専門家がコードなしで独自の翻訳モデルを構築することができる。

Translation API の Basic では、Google の NMT を用いて 100 以上の言語を翻訳することが可能である。また、Advanced では、これらにカスタマイズ機能を搭載し、分野やコンテキスト固有の用語やフレーズ、フォーマット済の文書翻訳に対応する。

「Media Translation」は、音声データから直接、コンテンツとアプリケーションにリアルタイムの音声翻訳を提供しているサービスである。

(2) 翻訳エンジン・アルゴリズム・訳質

翻訳エンジン及びアルゴリズムに関する詳細情報は公開されていない。一方で、2016 年 9 月から翻訳アルゴリズムを変更し、ニューラルネットワークを用いた「Google Neural Machine Translation (GNMT)」を導入し、同年 11 月からは、Zero-Shot Translation⁵を導入した。これは、AI は翻訳の際に用いる「中間言語」のようなものを独自で内部に作り、まだ学習していないはずの言語の組み合わせでも一定の翻訳を行う能力であり、それ以前の翻訳では、英語以外のある言語から別の言語に翻訳する場合には一度英語を介して翻訳を行う仕組み (中間言語としての英語) になっていたが、Zero-Shot Translation が行われ、英語を介さない翻訳が実現した。

⁴ Cloud Translation について <<https://cloud.google.com/translate/docs/overview?hl=ja>>. *2024 年 3 月 12 日調べ。

⁵ <https://blog.research.google/2016/11/zero-shot-translation-with-googles.html>

また、前述の「AutoML Translation」の機械学習では、データを使用してアルゴリズムをトレーニングする必要があると示されている。但し、アルゴリズムの詳細やトレーニング方法は、問題空間に応じて変わるとした。

モデルの評価において、「AutoML Translation」では、BLEU (Bilingual Evaluation Understudy) スコアで評価している。一方で、BLUE スコア値の解釈に関して、異なるコーパス間や言語間で BLEU 値を比較することは推奨されない、としている⁶。

(3) 対象言語

対象言語は、134 言語に対応しており、国内の地域言語への対応が行われている。本調査の対象言語に対しては、言語の両方向とも翻訳に対応している。

表 3 本調査対象言語への対応状況 (Google Translate)

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | ○ | ○ |
| ポルトガル語 | ○ | ○ |
| イタリア語 | ○ | ○ |
| インドネシア語 | ○ | ○ |
| マレーシア語 | ○ | ○ |
| ベトナム語 | ○ | ○ |
| タイ語 | ○ | ○ |
| アラビア語 | ○ | ○ |
| ヒンディー語 | ○ | ○ |
| トルコ語 | ○ | ○ |

(4) 追加機能等 (追加学習)

Google Translate の追加学習機能では、AutoML テクノロジーを使用して、50 を超える言語ペアでカスタムモデルを構築することができる。Google の翻訳サービスでは、業界用語や分野固有の用語に対応できるようにカスタマイズ可能で、専門文書、商品説明書、ソーシャルコンテンツの翻訳において、コンテキストと意味を正しくすることができる。

「AutoML Translation」では、ソース言語とターゲット言語で一致する文のペアを使用してカスタムモデルのトレーニングを行うことができる。また、モデルのトレーニング後、テストセットの項目を使用して、BLEU スコアを使用して、新しいモデルの品質と精度を評価している同機能では、モデルの品質を表現する。

(5) その他 (料金等)

Google Translate は、2006 年から統計的機械翻訳 (SMT) を提供し、2015 年からニューラ

⁶ Google Cloud-Cloud Translation 「モデルを評価する」
<https://cloud.google.com/translate/docs/advanced/automl-evaluate?hl=ja>

ル機械翻訳を導入した。

Google Translate の料金は、クラウドサービスの利用料と関連する。Cloud Translation (Basic) の料金は、最初の 50 万文字 (1 か月当たり) までは無料であるが、50 万～10 億文字 (1 か月当たり) で\$20/100 万文字である。ただし、10 億文字以上となった場合は、割引料金があるとされる。実際の利用にあたっては、これらの利用料に加え、Google Cloud の利用料がかかる。また、Advanced では、Translation Document が DOCX、PPT、PDF 形式のみ 1 ページ当たり \$0.08 である。なお、カスタムモデルは別料金 (例：100 万文字=\$80+ カスタムモデルのトレーニング費用 (時間単価：\$45)) となっている。

翻訳サービスの中には、Media Translation サービスがある。この翻訳サービスは、音声認識を使用可能な機械学習モデルを用いたもので、料金は最初 60 分の標準タイプは無料である。60 分～100 万分以降では、標準モデルで\$0.068/分であり、拡張モデルで\$0.084/分 (1 か月あたりの利用は 100 万分まで) である。これらのサービスの利用には、前述のとおり、Google Cloud 利用料が生じる。

2.2.3. 事例 2 : Microsoft Translator⁷

(1) サービスの概要

Microsoft Translator は、Microsoft Cognitive Service の一部。企業向けクラウドサービスも提供している。翻訳製品は、2007 年から使用されるテキスト翻訳 (2011 年からは顧客向けの API を提供)、ドキュメント翻訳、翻訳カスタマイズを提供している。翻訳カスタマイズについては、Custom Translator を利用することで、企業、アプリ開発者、言語サービスプロバイダは、自社のビジネスや業界で使用されている専門用語を理解したニューラル翻訳システムを構築することができる。

Microsoft の AI スーパーコンピュータ、特に Microsoft Cognitive Toolkit のスケールとパワーを活用し、Translator は現在、ニューラルネットワーク (LSTM) の翻訳品質向上の新しい翻訳を提供している。

(2) 翻訳エンジン・アルゴリズム・訳質

機械翻訳エンジン、アルゴリズムに関する詳細情報は、公開されていない。機械翻訳エンジンのアルゴリズムは、Transformer 型が用いられており、いくつかの機械翻訳ベンチマークで一貫した BLEU の改善 (1 BLEU ポイント以上) を果たした、としている⁸。

⁷ Microsoft translator <<https://www.microsoft.com/ja-jp/translator/>>

⁸ Bei Li, Yi Jing, Xu Tan, Zhen Xing, Tong Xiao, Jingbo Zhu (2023) 「TranSFormer: Slow-Fast Transformer for Machine Translation」 <<https://www.microsoft.com/en-us/research/publication/transformer-slow-fast-transformer-for-machine-translation/>>

(3) 対象言語

対象言語は、100 言語以上であり、近年は、ローカル言語への対応を図る。2023 年 4 月には、インドのコンカニ語、マイティリ語、シンディー語を追加。対応言語数は、2024 年 1 月時点で 128 言語である。

表 4 本調査対象言語への対応状況 (Microsoft Translator)

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | ○ | ○ |
| ポルトガル語 | ○ | ○ |
| イタリア語 | ○ | ○ |
| インドネシア語 | ○ | ○ |
| マレー語 | ○ | ○ |
| ベトナム語 | ○ | ○ |
| タイ語 | ○ | ○ |
| アラビア語 | ○ | ○ |
| ヒンディー語 | ○ | ○ |
| トルコ語 | ○ | ○ |

(4) 追加機能等 (追加学習)

【Microsoft Translator Hub】

特定のデータセットを学習・展開させることで、任意の分野に特化した機械翻訳を実現している。チューニングにあたっては、BLEU スコアを併用しつつ、Microsoft TrueText を用いたチューニングを実施している⁹。人間の翻訳品質は約 80%、通常の機械翻訳は 30～40%。チューニングすると 40～60%まで向上するとされる (Microsoft Tech Summit 2017)。

【ニューラル辞書】

ニューラル辞書は、文脈全体に適合する方法で用語、語句を翻訳し、流暢で首尾一貫した翻訳を実現する。ダイナミック辞書 (動的辞書)、フレーズ辞書を拡張したもの。ニューラル辞書の使用により翻訳品質は全体的に改善する。利用可能な言語は、英語⇄中国語 (簡体字)、フランス語、ドイツ語、イタリア語、日本語、韓国語、ポーランド語、ロシア語、スペイン語、スウェーデン語であり、Custom Translator を使用している顧客が対象である (2023 年 12 月 6 日現在)¹⁰。

(5) その他 (料金等)

Microsoft Translator の料金は、無料版の場合、標準翻訳 (テキスト翻訳：言語検出、二か

⁹ Microsoft 「BLEU スコアとは」〈<https://learn.microsoft.com/ja-jp/azure/ai-services/translator/custom-translator/concepts/bleu-score>〉

¹⁰ Microsoft 「ニューラル辞書」〈<https://learn.microsoft.com/ja-jp/azure/ai-services/translator/neural-dictionary>〉

国語辞書、翻訳) とカスタム翻訳 (トレーニング) がある。標準翻訳とカスタム翻訳のトレーニング料金の任意の組み合わせで毎月 200 万文字は無料である。また、従量課金制の場合、標準翻訳では、標準翻訳 100 万文字あたり\$10 で、ドキュメント翻訳は 100 万文字あたり\$15 である。また、カスタム翻訳では、テキスト翻訳、ドキュメント翻訳がそれぞれ 100 万文字あたり \$40 であり、トレーニングについてはトレーニングデータのソースおよびターゲット 100 万文字あたり\$10 であり、トレーニングあたり最大\$30 である。カスタムモデルホスティングについては、ホストされたカスタム翻訳モデルあたり毎月\$10 (リージョン単位) となっている¹¹。

2.2.4. 事例 3 : DeepL¹²

(1) サービスの概要

2016 年に、Jaroslaw Kutylowski 博士が創業し、2017 年 8 月にドイツ・ケルンで DeepL SE 事業を展開。31 言語をサポート。DeepL Proha, ビジネス向けのサービスであり、自然的な訳文を提供。世界中で 20000 社以上が活用し、翻訳精度は競合他社 (Google、Amazon、Microsoft) の 3 倍以上の精度で翻訳可能である。

法務チームで使用可能なセキュアな AI 翻訳機能を、DeepL Pro では提供している。世界各地の弁護士や法律事務所、政府機関に採用され、EU の公文書を使って学習した AI が、特に法律関連のテキストで高精度な訳文を出力している。また、ドイツ並びに国際的な個人情報保護法を完全に遵守している。企業向けには、個別に NDA を作成し、業務上取扱う機密内容を確実に保護する。

サービスは、無料版 (文字数上限 5000 文字) と、DeepL Pro 版がある。

(2) 翻訳エンジン・アルゴリズム・訳質

DeepL のニューラルネットワークにも、アテンション機構など Transformer のアーキテクチャを一部採用している。一般に公開されている Transformer のアーキテクチャとは異なり、ネットワークのトポロジーに違いがある¹³。

2020 年と 2021 年、社外のプロの翻訳者に、翻訳システムの名前を伏せたうえで訳文を示し、それぞれの品質を評価している。評価には、様々な分野のテキストから 119 段落を抽出して使用した。

¹¹ Microsoft 「Azure AI 翻訳の価格」〈<https://azure.microsoft.com/ja-jp/pricing/details/cognitive-services/translator/>〉

¹² DeepL 〈<https://www.deepl.com/ja/translator>〉

¹³ DeepL (2021) 「DeepL の仕組み」, 2021 年 11 月 1 日. 〈<https://www.deepl.com/ja/blog/how-does-deepl-work>〉

(3) 対象言語

翻訳対象言語は、英語、日本語、中国語、ブルガリア語、チェコ語、デンマーク語、ドイツ語、ギリシャ語、スペイン語、エストニア語、フィンランド語、フランス語、ハンガリー語、インドネシア語、イタリア語、韓国語、リトアニア語、ラトビア語、ノルウェー語、オランダ語、ポーランド語、ポルトガル語、ルーマニア語、ロシア語、スロバキア語、スロベニア語、スウェーデン語、トルコ語、ウクライナ語、アラビア語の 30 言語に対応している。

本調査の対象言語においては、スペイン語、ポルトガル語、イタリア語、インドネシア語、アラビア語の 5 つの言語に対して、翻訳元言語・翻訳先言語として対応している。

表 5 本調査対象言語への対応状況 (DeepL)

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | ○ | ○ |
| ポルトガル語 | ○ | ○ |
| イタリア語 | ○ | ○ |
| インドネシア語 | ○ | ○ |
| マレー語 | × | × |
| ベトナム語 | × | × |
| タイ語 | × | × |
| アラビア語 | ○ | ○ |
| ヒンディー語 | × | × |
| トルコ語 | ○ | ○ |

(4) 追加機能等 (追加学習)

【語調を指定】

ターゲット読者層に合わせて、瞬時に訳文を書き換えることができる。具体的には、敬称と親称の使い分けであり、語調を敬称または親称に切り替えて、訳文の人称代名詞やこれに関連する箇所を自動変換できる。本機能は、イタリア語、オランダ語、スペイン語、ドイツ語、フランス語、ポーランド語、ポルトガル語、ポルトガル語 (ブラジル)、ロシア語で利用できる。

【オリジナルの用語集の作成】

文脈に即した語句を自動で適用するように設定して、訳文をカスタマイズすることができる。具体的には、重要語句の訳し方についてルールを設定し、常にルールに基づき訳出することで、編集に要する時間を大幅に減らすことができる。DeepL Pro のすべてのプランで、上限なしに用語ペアを登録できる。また、DeepL Pro の Advanced と Ultimate のプランでは、複数の用語集を作成したり、手持ちの用語集をアップロードしたりできる。用語集に使える言語の組み合わせは、英語⇄ドイツ、英語⇄フランス語、英語⇄スペイン語である。

(5) その他 (料金等)

DeepL の開発者向けの料金は、無料版で提供されるサービスは、DeepL REST API へのアクセス等、すべての機能を利用可能であるものの、1 か月に 500,000 文字まで翻訳、1,000 個の用語集 (一部言語のみ) 等の制限がある。DeepL API Pro 版は、月額 630 円で提供され、無料版からはセキュリティ対策が追加されるとともに、翻訳のテキスト量の制限はない。

2.2.5. 事例 4 : SAP Translation Hub

(1) サービスの概要

SAP が自社製品のユーザー向けに開発した、機械翻訳クラウドプラットフォームである。SAP Translation Hub では、SAP が承認した翻訳と用語のリポジトリ、および機械翻訳を使用して、UI テキストおよび文書を翻訳する。

SAP Translation Hub の機能は、UI テキスト翻訳、API の利用、ABAP 拡張の翻訳、翻訳プロジェクトの管理、統合シナリオの使用、独自の言語データの統合、MT を使用した文書の翻訳からなる。上記の機能の内、UI テキスト翻訳は、多言語テキストリポジトリ (MLTR)、機械翻訳ソリューション、または独自の言語データを含む会社の多言語テキストリポジトリ (会社 MLTR) の複数の翻訳プロバイダを使用して、UI テキストを翻訳するものである。また、MT を使用した文書の翻訳は、SAP Translation Hub の Document Translation サービスでは、ユーザーが機械翻訳 (MT) を使用して文書を翻訳することができる。

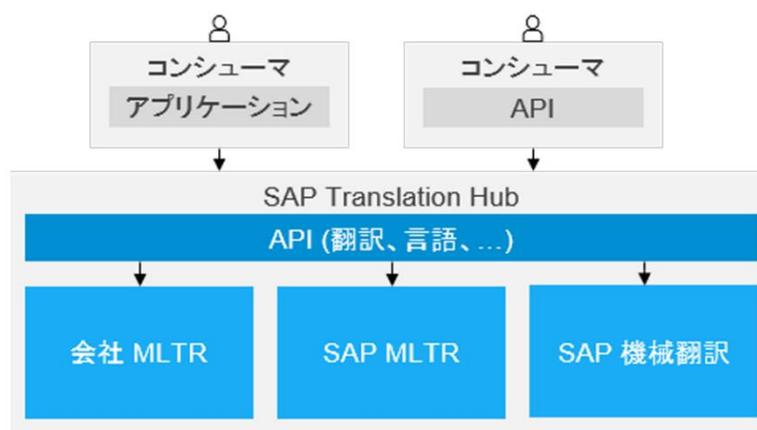


図 3 SAP Translation Hub の翻訳ワークフロー

出所：SAP Translation Hub ユーザーガイド

https://help.sap.com/doc/392c858a3cc8489d993d1abc060b642c/Cloud/ja-JP/SAP_Translation_Hub_JA.pdf

SAP Translation Hub は、以下の翻訳プロバイダで構成される。

- SAP アプリケーションからの多言語テキストのデータベース：多言語テキストは、多言語テキストリポジトリ（MLTR）と呼ばれる SAP HANA データベースに保存される。
- SAP のトレーニング済みエンジンからの機械翻訳：SAP が認定した言語エキスパートによって検証された多数の翻訳についてトレーニング
- オプション: ユーザー独自の言語データを SAP Translation Hub にアップロードすると、優先される翻訳プロバイダとしてテキストの独自のリポジトリを使用できる。このリポジトリは、会社 MLTR と呼ばれる。

SAP Translation Hub では、ドメインの選択ができる。ドメインは、分析、財務会計、一般、業種（自動車、航空宇宙・防衛、保険等の SAP 業種別ソリューション）、ロジスティクス、人材管理、販売、テクノロジーからなる。

SAP 機械翻訳（SAP MT）は、テキストを別の言語に自動的に翻訳するソフトウェアである。SAP MT では、SAP が認証した言語エキスパートにより検証された多数の翻訳が記憶されている。SAP MT は、SAP に関連するコンテンツの翻訳に特化し、オープンソースの MT システムとは異なり、ドメインの適合と用語の範囲に優れている。

(2) 翻訳エンジン・アルゴリズム・訳質

SAP Translation Hub のニューラル機械翻訳エンジンを用いており、アルゴリズムは、ニューラルネットワークによる高精度のアルゴリズム（Transformer）を用いている¹⁴。

翻訳品質については、Transformer アーキテクチャを使用することで、SAP の機械翻訳の翻訳品質は大幅に向上した。BLEU は、14 の言語ペア¹⁵について、平均で+5.2 BLEU%が増加した¹⁶。また、別の機械翻訳サービスとの出力比較では、SAP Translation Hub の出力の品質が一番良いと評価された（セグメント数は平均で 65 セグメントで、3 つのエンジンのうち最も品質が高いのは SAP Translation Hub と評価された）¹⁷。

(3) 対象言語

対象言語は、45 言語であり、本調査の対象言語では、アラビア語、ヒンディー語、インドネシア語、イタリア語、マレー語、ポルトガル語（ブラジル／ポルトガル）、スペイン語、タイ語、トルコ語、ベトナム語に対応している。これらの言語は、翻訳元言語、翻訳先言語に使用することができる。英語⇄対象言語のペアの場合は、翻訳指数スコアが提供

¹⁴ SAP Community 「Machine translation by SAP: our latest updates behind the scenes」, Technology Blogs by SAP 〈<https://community.sap.com/t5/technology-blogs-by-sap/machine-translation-by-sap-our-latest-updates-behind-the-scenes/ba-p/13470345>〉

¹⁵ BLEU 値が増加した 14 の言語ペア：英語⇄ドイツ語、スペイン語、フランス語、日本語、ブラジルポルトガル語、ロシア語、中国語

¹⁶ 前掲：SAP Community

¹⁷ 川村インターナショナル「SAP 翻訳に特化した SAP Translation Hub とは」.
〈https://www.k-intl.co.jp/blog/B_190719A〉

される。

表 6 本調査対象言語への対応状況 (SAP Translation Hub)

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | ○ | ○ |
| ポルトガル語 | ○ | ○ |
| イタリア語 | ○ | ○ |
| インドネシア語 | ○ | ○ |
| マレー語 | ○ | ○ |
| ベトナム語 | ○ | ○ |
| タイ語 | ○ | ○ |
| アラビア語 | ○ | ○ |
| ヒンディー語 | ○ | ○ |
| トルコ語 | ○ | ○ |

(4) 追加機能等 (追加学習)

SAP Translation Hub は、SAP ネットワークに組み込まれているシステムであり、翻訳エンジンのトレーニングに必要な言語資源を随時得ることができる環境である。

(5) その他 (料金等)

「XMAT for SAP Translation Hub」は、SAP 社が自社製品ユーザー向けに開発した機械翻訳サービスである。国内では、川村インターナショナル社がユーザー、1 か月あたり 5500 円から提供している。

2.2.6. 事例 5 : SYSTRAN (シストラン)

(1) サービスの概要

1968 年に世界初の自動翻訳システムを商用開発し、自動翻訳に特化した研究開発企業として、機械翻訳サービスを 50 年以上、各国政府、国際機関等に提供している。導入機関の中には、OECD、米国連邦政府インテリジェンス・コミュニティ、仏・国防省、NATO/OTAN 等の機関も含まれる。

SYSTRAN は、法務・特許、金融・財務、製造、医療・製薬、IT・ゲーム等の利用シーンに対応することができる。

(2) 翻訳エンジン・アルゴリズム・訳質

翻訳エンジンは、PNMT™ (Pure Neural™ Machine Translation)。カスタマイズ可能で、法律、自動車産業、IT、観光等に至るまで、各顧客の特定のコンテキストに適応した機械翻訳を提供 (専門分野のエンジン：製造、金融、医療・製薬、法務、IT、テクノロジー、教育、

農業、特許、ビジネス、対話、汎用エンジン)。

SYSTRAN は、自然言語処理の研究における最新のブレイクスルーと人間の専門知識を組み合わせて、モデルを慎重に訓練及び評価している。また、システムは AI 翻訳に由来するニューラルネットワークアーキテクチャであるトランスフォーマーモデルを使用している¹⁸。また、独自のデータや翻訳メモリを使用して、トレーニングされ、独自にカスタマイズされた翻訳モデルを作成し、翻訳の生産性を大幅に向上させることができる¹⁹。

訳質については、国際的な評価機関『Intento』と『TAUS』によって行われた、国内外の 29 の主要な翻訳エンジンを対象としたレポートで、SYSTRAN の翻訳エンジンが高い評価を獲得した。専門家 4 人が、学習済みの専門分野の AI エンジン (Neural MT) と従来の確率エンジン (SMT) をブランド評価した結果、AI エンジンの翻訳精度は、専門家による翻訳精度にかなり近いという結果であった。

(3) 対象言語

対象言語は、日本語、英語、中国語<簡・繁>、韓国語、ロシア語、タイ語、インドネシア語、ベトナム語、マレーシア語、フィリピン語、ミャンマー語、ヒンディー語、ウルドゥー語、アラビア語、ペルシャ語、トルコ語、ヘブライ語、ドイツ語、スペイン語、フランス語、ポルトガル語、イタリア語、ギリシャ語、オランダ語、ノルウェー語、スウェーデン語、デンマーク語、フィンランド語、ポーランド語の 28 言語に対応している。

本調査の対象言語については、タイ語、インドネシア語、ベトナム語、マレーシア語、ヒンディー語、アラビア語、トルコ語、スペイン語、ポルトガル語、イタリア語等、幅広い言語をカバーしている。

表 7 本調査対象言語への対応状況 (SYSTRAN)

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | ○ | ○ |
| ポルトガル語 | ○ | ○ |
| イタリア語 | ○ | ○ |
| インドネシア語 | ○ | ○ |
| マレー語 | ○ | ○ |
| ベトナム語 | ○ | ○ |
| タイ語 | ○ | ○ |
| アラビア語 | × | × |
| ヒンディー語 | ○ | ○ |
| トルコ語 | ○ | ○ |

¹⁸ SYSTRAN Translate Server <<https://www.systransoft.com/jp/translation-products/translate-server/>>

¹⁹ SYSTRAN model studio <<https://www.systransoft.com/jp/translation-products/systran-model-studio/>>

(4) 追加機能等（追加学習）

【SYSTRAN Pure NeuralR Server】

SYSTRAN Pure NeuralR Server は、原文を最初から人手で翻訳するよりも、AI の翻訳結果を少しポストエディット（手直し）することで、翻訳プロセス全体をスピードアップすることができる。同時に翻訳知識として用語集や翻訳メモリに保存することで、翻訳済みの文を再度翻訳する必要がなくなり、翻訳プロセスがさらにスピードアップできる。

【SYSTRAN Model Studio】

上記で増え続ける翻訳知識を AI が追加学習する。さらに専門化した AI 翻訳エンジンが、翻訳済み以外にも文も正確に翻訳を実施し、ポストエディット作業が大幅に削減され、翻訳プロセスのさらなるスピードアップにつながる。

(5) その他（料金等）

SYSTRAN の機械翻訳サービス形態は、ホスティング型、オンプレミス型、専用サーバー型に分類できる。

2.2.7. 事例 6：みらい翻訳²⁰

(1) サービスの概要

Mirai Translator は、企業内の文書翻訳の AI 自動翻訳ソリューションであり、ユーザー辞書機能を利用したテキストやファイルの翻訳を、グループ単位で活用できる。翻訳エンジンは、NICT の研究成果を活用したものである。

翻訳サービスの特徴は、分野専門モデルの搭載（契約書・規定等の法務関連の文書、決算短信・有価証券報告書等の財務関連文書の翻訳に適した法務・財務モデル、特許公報・明細書等の特許書類特有の言い回し・長文に対応した特許モデルが利用可能であることである。

サービスの展開にあたっては、国内クラウド機械翻訳で ISO27017 認証を取得し、ISO27001 (ISMS) とあわせ、これらの認証に準拠している。

2024 年 1 月 30 日からは、新サービスとして、「みらい翻訳 Plus」では、AI 翻訳技術 (NMT) と生成 AI (大規模言語モデル: LLM) を融合した翻訳サービスの提供を開始した。サービス面では、二次利用は行わず、翻訳完了とともにデータをサーバー上から削除している。テキスト翻訳のほか、PDF・PowerPoint・Word・Excel・TXT などのファイル翻訳機能を用意している。

²⁰ みらい翻訳 〈<https://miraitranslate.com/>〉

(2) 翻訳エンジン・アルゴリズム・訳質

翻訳エンジンは、国立研究開発法人情報通信研究機構（NICT）の研究成果を利用し、社内でエンジン開発しており、2019年には製薬企業との共同プロジェクトにより「製薬専門モデル」を、2020年にはアンダーソン・毛利・友常法律事務所と「法務・財務モデル」を開発した。特許文書の翻訳に特化したモデルは、NTTから特許コーパス（3億文）の提供を受けている²¹。

新サービスの「みらい翻訳 Plus」は、TOEIC960点レベルの自社翻訳エンジンに、生成AIが得意とする文章の変更や英文校正等の機能を加えたサービスを展開している。ビジネスコミュニケーションや経済系ニュース等を中心に翻訳精度が向上し、情報伝達と流暢さの観点で最大32%向上している²²。

(3) 対象言語

Mirai Translatorは、12言語に対応し、標準言語ペアは、「日本語⇄英語、中国語（簡体字／繁体字）」、「英語⇄中国語（簡体字／繁体字）」である。また、法務・財務モデル提供言語ペアは、「日本語⇄英語、中国語（簡体字／繁体字）」である。特許モデル提供言語ペアは、「日本語⇄英語」に留まる。それ以外に、多言語サービス提供言語ペアとして、「日本語⇄欧州系言語（ドイツ語、スペイン語、フランス語、ポルトガル語、ロシア語、イタリア語）、アジア系言語（韓国語、タイ語、ベトナム語、インドネシア語）」と「英語⇄欧州系言語（ドイツ語、スペイン語、フランス語、ポルトガル語、ロシア語、イタリア語）、アジア系言語（韓国語、タイ語、ベトナム語、インドネシア語）」がある²³。

また、みらい翻訳 Plusでは、日、英、中（簡）、中（繁）、韓、タイ、ベトナム、インドネシア、イタリア、スペイン、ドイツ、フランス、ポルトガル、ロシアの14言語に対応している。

表8 本調査対象言語への対応状況（Mirai Translator）

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | ○ | ○ |
| ポルトガル語 | ○ | ○ |
| イタリア語 | ○ | ○ |
| インドネシア語 | ○ | ○ |
| マレー語 | × | × |
| ベトナム語 | ○ | ○ |
| タイ語 | ○ | ○ |
| アラビア語 | × | × |
| ヒンディー語 | × | × |

²¹ みらい翻訳（2023）「特許専用 AI 自動翻訳モデル発売のお知らせ」, 2023年10月17日. <<https://miraitranslate.com/news/1936/>>

²² みらい翻訳「みらい翻訳 Plus」 <<https://plus.miraitranslate.com/>>

²³ Mirai Translator®サービス資料（みらい翻訳からの入手資料）

| | 日本語→当該言語 | 当該言語→日本語 |
|------|----------|----------|
| トルコ語 | × | × |

注) 特許翻訳エンジンは、日本語⇄英語のみ。

(4) 追加機能等 (追加学習)

【ユーザー辞書】

製品名や専門用語、特殊な社内用語など、表現の決まっている名詞をユーザー辞書機能で登録することで、翻訳結果をカスタマイズすることが可能である。

【翻訳メモリ】

定型的に繰り返される文章で、翻訳結果も固定したい場合、翻訳メモリとして文章を登録し翻訳結果に反映することが可能です。メールの定型文や、マニュアル等の翻訳に活用することができる。

(5) その他 (料金等)

Mirai Translator は、2020年6月1日に提供開始し、法務・財務専用エンジンは2021年1月27日に、特許翻訳特化型エンジンは2023年10月18日に提供を開始した。また、みらい翻訳 Plus は、2024年1月30日に提供開始した新サービスである。

料金形態は、従量型「ワード数定額プラン」が年額：120,000円（税抜）からであり、12万ワードの翻訳が可能である。また、定額「ベーシックプラン」は、月額：80,000円（税抜）からのプランで10ライセンスまで利用が可能である。多言語翻訳、法務・財務モデル利用はオプションで別料金となっている²⁴。

みらい翻訳 Plus については、Free版（無料）と、Starter版（500円/月）の2つのプランがある。Free版は、1か月1ファイルの翻訳に留まる一方で、Starterは、1か月に10ファイルまで翻訳ができるとしている。また、入力文字数の無制限化も予定している。

2.2.8. 事例7：みんなの自動翻訳

(1) 概要 (サービスの提供形態)

「みんなの自動翻訳」は、国立研究開発法人情報通信研究機構（NICT）が開発した自動翻訳技術であり、AI技術で多用される深い階層構造を持つネットワークを用いた自動翻訳技術（AI自動翻訳）である。「みんなの自動翻訳」は、TexTraとして、WordやPowerPointのファイルを直接翻訳する機能を搭載している。

法人向けの商用版（川村インターナショナル社が提供）では、汎用的な翻訳から、文書

²⁴ みらい翻訳「Mirai Translator 特許モデルのご提案（資料）」（みらい翻訳からの入手資料）

の内容に応じて、特許、特許審決等、金融（IR/適時開示）、金融サービス、法令契約、サイエンス、ニュースと専門分野の機械翻訳サービスを提供している。2022年2月より、バージョン7.3を提供している。

(2) 翻訳エンジン・アルゴリズム・訳質

NICTのニューラル機械翻訳エンジンを用いているため、アルゴリズムは、ニューラルネットワークによる高精度のアルゴリズム（Transformer）を用いている^{25,26}

訳質について、商用版の提供事業者は、みんなの自動翻訳 Ver.3.0から Ver.4.0の提供にあたり、自社の翻訳者による評価として訳質の評価を実施した。具体的には、内容が正確で自然な訳文になっているか、内容はあっているが修正を加える必要があるか、訳揺れ・訳抜け等のエラーがあるかの観点で評価した。

(3) 対象言語

対象言語については、非商用版の汎用 NT は、21 言語（日本語⇄ポルトガル語、中国語（繁体字）、中国語（簡体字）、ベトナム語、ウクライナ語、タイ語、ロシア語、ネパール語、ミャンマー語、モンゴル語、クメール語、イタリア語、インドネシア語、ヒンディ語、フランス語、フィリピン語（タガログ語）、英語、スペイン語、韓国語、ドイツ語、アラビア語）に対応し、対話 NT（音声翻訳エンジン）やその他のエンジンを加えると、106 言語に対応する²⁷。高い翻訳精度や強固なセキュリティを備えた商用版は、日本語に対し、英語、簡体字、繁体字、韓国語、ドイツ語、フランス語、スペイン語、インドネシア語、ミャンマー語、ポルトガル語、タイ語、ベトナム語、アラビア語、イタリア語、ロシア語、フィリピン語、クメール語、モンゴル語、ネパール語等のへの翻訳が可能である。また、特許特化型の商用版エンジンの対象言語は、日本語に対し、英語、簡体字、繁体字、韓国語、ドイツ語、フランス語の6言語への翻訳が可能である。

表9 本調査対象言語への対応状況（みんなの自動翻訳）

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | ○ | ○ |
| ポルトガル語 | ○ | ○ |
| イタリア語 | ○ | ○ |
| インドネシア語 | ○ | ○ |

²⁵ 内山将夫（2021）「自動翻訳技術の概要：なにができるか／できるようになってきているか」, 特許情報シンポジウム, 2021年2月26日. <<https://www2.nict.go.jp/astrec-att/member/mutiyama/pdf/2021-patent-sympo.pdf>>

²⁶ 内山将夫（2019）「翻訳バンクとアダプテーションによるNMT超高精度化」, 特許・情報フェア&コンファレンス. <<https://www2.nict.go.jp/astrec-att/member/mutiyama/pdf/pifc-2019-Nov-6-7.pdf>>

²⁷ みんなの自動翻訳 <https://mt-auto-minhon-mlt.ucri.jgn-x.jp/content/mt/?q_tab=0&q_offset=0>

| | | |
|--------|---|---|
| マレー語 | × | × |
| ベトナム語 | ○ | ○ |
| タイ語 | ○ | ○ |
| アラビア語 | ○ | ○ |
| ヒンディー語 | ○ | ○ |
| トルコ語 | ○ | ○ |

(4) 追加機能等（追加学習）

「みんなの自動翻訳」非商用版では、自動翻訳結果の修正、対訳集の登録が可能であり、専用の自動翻訳エンジンを登録・訓練することが可能である。また、用語集で訳語を指定しカスタムした自動翻訳を構築することができる。

また、カスタム自動翻訳では、複数の自動翻訳エンジン（汎用、特許、対話、FSA、金融、法令契約、汎用+実験用・技術移転不可）を同時に使うことが可能である。

(5) その他（料金等）

みんなの自動翻訳は、統計的機械翻訳（SMT）を2014年6月に提供を開始し、2017年6月にニューラル機械翻訳（NMT）を提供した。

2.2.9. 事例8：T-400²⁸

(1) サービスの概要

T-400は、株式会社ロゼッタが提供する機械翻訳サービスである。T-400は、産業機械・電子部品・製薬分野などのガイドライン・公的文書を大量に学習させた翻訳エンジンと専門分野のデータベースを用いてAI自動翻訳を行う。データベースは、専門分野データベース（2000分野のDB）、企業別データベース（契約企業毎に割振られる専用データベース）²⁹の2つからなる。専門分野のデータベースは、「医薬・化学・環境」、「IT・機械・電気電子」、「ビジネス専門」、「人文・社会」に大別される³⁰。これらのデータベースには、

²⁸ AI自動翻訳「T-400」

〈https://www.rozetta.jp/t400/?gad_source=1&gclid=Cj0KCKQjwwMqvBhCtARIsAIXsZpYg63fS58iN_d4ozbchIfp69S0Bao9hAtSrsxhuTTvAOuSCD-O3xeUaAr3fEALw_wcB〉

²⁹ 企業別データベースは、ユーザーの社内にある英語・日本語の文書を企業別データベースに登録することで、社内表現や言い回しをAIが学習する。

³⁰ 専門分野データベースについて、「医薬・化学・環境」区分は、医学、ライフサイエンス、薬事、医療機器、化学、環境、食品、農林水産、特許（医学・バイオ・化学）、医薬・化学・環境（その他）で構成される。また、「IT・機械・電気電子」区分は、IT・通信、機械、電気電子、金属、土木・建築・鉱山、エネルギー、プラント、物流・運輸・交通、写真・印刷技術、日用品・家具・雑貨、生産管理、IT・機械・電気電子（規制）、自然科学、特許（IT・機械・電気電子）、IT・機械・電気電子（その他）等で構成される。「ビジネス専門」区分は、会社規定、不動産、法務、特許・知財、金融・財務・経理、マーケティング、商業・貿易、HP・会社案内・広報・PR、ビジネス（その他）等で構成される。それ以外に「人文・社会」区分がある。

各分野には国内外の公的文書・ガイドライン等が収録されている。

T-400 の機械翻訳サービスでは、プロの翻訳者の意見を取り入れて編集機能を搭載しているほか、類似文書の訳語のサジェスト機能、大規模言語モデル（LLM）による生成 AI 翻訳にも対応している。生成 AI 翻訳は、従来の NMT で発生しやすい訳抜け問題を抑え、文脈や背景までを踏まえた自然な翻訳を実現した。サービスの提供にあたっては、サーバーは国内に設置され、ISMS 認証取得、暗号化通信にも対応している。

(2) 翻訳エンジン・アルゴリズム・訳質

翻訳エンジン、アルゴリズムに関する詳細情報は、公開されていない。T-400 の翻訳エンジンは、産業機械・電子部品・製薬分野などのガイドライン・公的文書を大量に学習させた自社開発の翻訳エンジンと、カスタムエンジン（世界で展開されている汎用翻訳エンジン）の 2 つを用いている。自社開発の翻訳エンジンは、23 言語（後述）のみ対応しており、23 言語以外の言語は、カスタムエンジンと T-400 のデータベースとの組合せでサービスを提供している。

翻訳精度については、医薬、金融、化学、法務等の専門文書の英日翻訳において、「最大 95%＝プロ翻訳レベルの正確さ」の翻訳³¹が可能であり、分野毎のチューニングにより、複雑な長文も正確で滑らかな訳文を実現している。なお、翻訳精度 95%以上の分野は「医学」「化学」「法務」「IT・通信」「金融・財務・経理」分野の英日翻訳であり、その他の産業翻訳の主要専門分野（「薬事」「機械」「電気電子」「特許等」等）は 90%以上の精度である³²。

(3) 対象言語

対象言語は、100 言語以上であるが、前述の自社開発の翻訳エンジンに搭載される言語は、次の 23 言語である（日本語⇄英語、中国語〈簡体字、繁体字〉、韓国語、アラビア語、ベンガル語、ドイツ語、ギリシャ語、スペイン語、ペルシャ語、フィンランド語、フランス語、ヒンディー語、インドネシア語、イタリア語、マレー語、ポルトガル語、ロシア語、スウェーデン語、タイ語、タガログ語、トルコ語、ベトナム語）。

表 10 本調査対象言語への対応状況（T-400）

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | ○ | ○ |
| ポルトガル語 | ○ | ○ |
| イタリア語 | ○ | ○ |
| インドネシア語 | ○ | ○ |
| マレー語 | ○ | ○ |
| ベトナム語 | ○ | ○ |

³¹ ロゼッタ社の人目評価 〈<https://prtimes.jp/main/html/rd/p/000000595.000006279.html>〉

³² T-400（ティーフォーオーオー）AI 自動翻訳 〈<https://www.softbank.jp/biz/ai/t-400/>〉

| | | |
|--------|---|---|
| タイ語 | ○ | ○ |
| アラビア語 | ○ | ○ |
| ヒンディー語 | ○ | ○ |
| トルコ語 | ○ | ○ |

(4) 追加機能等（追加学習）

T-400 では、専門分野ごとに AI を使って機械翻訳エンジンを学習させている。また、2023 年 6 月 7 日には、大規模言語モデル（LLM）をベースとした翻訳エンジンを搭載した。T-400 では、NMT 翻訳エンジンにより高精度かつ高速の自動翻訳を提供するとともに、生成 AI に基づくエンジンでは、文意を推測しながら訳文の作成・推敲を可能とした。

(5) その他（料金等）

T-400 の最初に提供を開始した時期は、2017 年 11 月である。2018 年 6 月には Version 2 を提供し、顧客別カスタマイズ、専門分野別〈90 種類以上の言語対応〉、翻訳エンジンの品質評価手法を確立（独自の QE 手法）した。2019 年 3 月の Version 3 では、AI による分野の自動判定機能を開始した。これらを踏まえ、2023 年 6 月に LLM ベースの生成 AI 翻訳エンジンの実装を開始した。

2.3. 特許情報に特化した機械翻訳エンジン

2.3.1. 各国知財庁の機械翻訳の提供状況

(1) 機械翻訳の提供有無

本項目では、日本国特許庁（JPO）、米国特許商標庁（USPTO）、欧州特許庁（EPO）、国家知識産権局（CNIPA）、韓国特許庁（KIPO）、英国知的財産庁（UKIPO）、世界知的所有権機関（WIPO）、欧州連合知的財産庁（EUIPO）を対象に、特許情報に特化した機械翻訳機能の提供状況について把握した。

機械翻訳機能を提供している知財庁は、日本国特許庁、欧州特許庁、中国国家知識産権局、韓国特許庁、世界知的所有権機関の 5 機関であった。

表 11 各国知財庁の機械翻訳の提供状況

| 知財庁 | 特許機械翻訳提供状況 | 機械翻訳エンジン |
|------------------|------------|-------------------------------------|
| 日本国特許庁（JPO） | 提供あり | NICT 開発エンジン（対訳コーパス・辞書によるカスタマイズ等を含む） |
| 米国特許商標庁（USPTO） | 提供なし | — |
| 欧州特許庁（EPO） | 提供あり | 独自開発エンジン（Google との協力） |
| 中国国家知識産権局（CNIPA） | 提供あり | 独自開発エンジン |
| 韓国特許庁（KIPO） | 提供あり | 独自開発エンジン（WIPO との業務提携） |

| | | |
|-------------------|------|-----------------------|
| 英国知的財産庁 (UKIPO) | 提供なし | — |
| 世界知的所有権機関 (WIPO) | 提供あり | 独自開発エンジン |
| 欧州連合知的財産庁 (EUIPO) | 提供なし | ※eSearch plus 意匠・商標検索 |

以下、各機関で提供している機械翻訳機能について示す。

(2) 日本国特許庁 (JPO) の機械翻訳機能の提供例³³

① サービスの概要

日本国特許庁では、特許情報プラットフォーム (J-PlatPat)、ワン・ポータル・ドシエ (OPD) の審査書類情報照会システムにおいて、日本の審査・審判書類が英語にリアルタイムで機械翻訳され、主に海外ユーザー向けに照会可能となっている他、中国・韓国公報が日本語に機械翻訳され、主に国内ユーザー向けに検索・照会可能となっている。

特許庁で運用している機械翻訳プラットフォームは、上記機械翻訳を実行するシステムであり、国立研究開発法人情報通信研究機構 (NICT) が開発したエンジンと、東芝デジタルソリューションズの自然言語処理技術を組み合わせ、従来の機械翻訳では難しく、正確で自然な機械翻訳を可能とした^{34,35}。具体的には、自然言語処理技術で、特許文献の明細書部分・請求項部分や定型部分を自動で切り分け、ニューラル機械翻訳 (NMT)、ルールベース機械翻訳 (RBMT)、統計的機械翻訳 (SMT) のそれぞれに適宜振り分け、誤訳が発生しにくいよう、入力文を整形する処理技術を実装している³⁶。なお、クラウド基盤は、Microsoft Azure を採用している。

また、外国特許情報サービス (FOPISER) では、外国特許庁から受領した公報を、Google 翻訳を利用しながら、検索・照会可能としている。

② 翻訳エンジン・アルゴリズム・訳質

翻訳エンジンは、NICT のニューラル機械翻訳エンジン (NMT) を技術移転し、RBMT (ルールベース機械翻訳)、SMT (統計的機械翻訳) エンジンを組み合わせたものである。2019年5月に日英翻訳機能をリリースし、2020年5月に中日、韓日翻訳機能を実装した。

NICT のニューラル機械翻訳エンジンを用いているため、アルゴリズムは、ニューラルネットワークによる高精度のアルゴリズム (Transformer) を用いている。

訳質は、日本公報の明細書の英訳の BLEU の平均スコアが 40 以上であり、中国、韓国公報の明細書の和訳はさらに、BLEU の平均スコアが最低でも 45、60 以上ある高品質の機械

³³ 特許庁 (2020) 「特許情報プラットフォームの機能改善について (中韓文献機械翻訳文の訳質向上)」。〈https://www.jpo.go.jp/support/j_platpat/kaizen20200518.html〉

³⁴ 中西聡 (2020) 「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—」 〈https://www.japio.or.jp/00yearbook/files/2020book/20_a_06.pdf〉

³⁵ Toshiba Clip (2019) 「あなたの知らない機械翻訳の世界 日本の知財戦略を支える翻訳システム」, 2019年8月21日。〈<https://www.toshiba-clip.com/detail/p=268>〉

³⁶ IT Leaders (2019) 「特許庁、「機械翻訳システム」の日英翻訳機能を活用」, 2019年6月3日。〈<https://it.impress.co.jp/articles/-/18002>〉

翻訳を提供している。

③対象言語

上記機械翻訳プラットフォームの対象言語は、日本語から英語、中国語から日本語、及び韓国語から日本語であり、本調査対象言語には対応していない。

④追加機能等

追加機能については、新語への対応を図っている。

⑤その他

機械翻訳機能の提供開始時期は、2000年3月に日英の統計的機械翻訳（SMT）を特許電子図書館で提供したが、その後、上記機械翻訳プラットフォームを構築し、2019年5月より、J-PlatPatにて、日英リアルタイム翻訳を提供し、2020年5月には、中日・韓日翻訳文を検索・照会可能な形で提供した。

(3) 欧州特許庁（EPO）の機械翻訳機能の提供例³⁷

①サービスの概要

EPOは、Googleと協力し、特許文書専用の機械翻訳サービス（Patent Translate）を提供している。同サービスは、32言語の特許文書を扱い、32言語⇔英語、フランス語、ドイツ語への翻訳が可能である。

Patent Translateは、EPOの特許検索サービス「Espacenet」³⁸にて提供している。具体的な手順は、「Espacenet」にて、特許文書を検索し、翻訳したい部分（要約、明細書、特許請求の範囲）を選択する。選択後、翻訳先の言語を選択肢、特許翻訳ボタンをクリックする形で利用することができる。

②翻訳エンジン・アルゴリズム・訳質

Googleの協力の下、特許文書専用の機械翻訳サービスを構築しているが、機械翻訳エンジンの詳細についての情報は公開されていない。アルゴリズムについては、Google翻訳機能は、ニューラル機械翻訳を提供しているため、ニューラルネットワークによる高精度のアルゴリズムを用いているものと推察される。

③対象言語

対象言語は、EPO各国の公用語（28言語）⇔英語、ドイツ語、フランス語への翻訳と、英語⇔日本語、中国語、韓国語、ロシア語に翻訳することができる。これら言語全体で、

³⁷ EPO「Patent Translate」〈<https://www.epo.org/en/searching-for-patents/helpful-resources/patent-translate>〉

³⁸ Espacenet 〈(<https://worldwide.espacenet.com/patent/help/patenttranslate>)〉

32 言語としている。本調査対象言語については、欧州地域のスペイン語、ポルトガル語、イタリア語には対応しているが、アジア、中東地域の言語には対応していない。

④追加機能等

追加機能については、新語への対応を図っている。

⑤その他

EPO での機械翻訳サービスは、2012 年から利用を開始している。機械翻訳サービスの提供にあたり、EPO と Google では、2011 年 3 月より特許の機械翻訳の改善のための協定を締結し、2012 年には、EPO のパラレル特許テキストを組み込み、英語⇄フランス語、ドイツ語、スペイン語、ポルトガル語、スウェーデン語の翻訳機能を提供するに至った³⁹。

また、これらサービスの利用料については、Espacenet、European Publication Server、Global Patent Index のユーザーは無料で利用できる。

(4) 中国国家知識産権局 (CNIPA) の機械翻訳機能の提供例

①サービスの概要

中国特許照会サービス、中外専利データベース・サービスは中国の提供機関から認証コードを受け取る必要があるため確認は不可である。また、特許検索・分析システムを利用するには、メールアドレスを用いた認証によるユーザー登録が必要となっている。

CNIPA の支局に位置する、中国特許情報センター (CNPAT)⁴⁰は、国有の大規模な特許情報サービス機関で、CNIPA および営利企業に特許データベースのサービスを提供している。CNPAT では、グローバル特許文書の翻訳等のサービスを提供している。CNPAT の製品・サービスの中に、專利之星検索平台 (CPRS : Patent Star Search Platform) があるが、日本からはアクセスが難しい。

②翻訳エンジン・アルゴリズム・訳質

中国特許情報センター (専利情報センター) は、2008 年に中国特許文献の中英機械翻訳システムを公表し、CNIPA のウェブサイトを通じて、中国特許情報のオンライン中英機械翻訳サービスを提供した。2019 年には、特許分野のニューラルネットワーク機械翻訳技術を導入した。2022 年に機械翻訳エンジンを独自設計・開発したとされる⁴¹。併せて、機械翻訳製品「ペセッタ」を出願している。

³⁹ Google (2012) 「Translating patents with the European Patent Office」, 2012 年 2 月 29 日。
〈<https://translate.googleblog.com/2012/02/translating-patents-with-european.html>〉

⁴⁰ CNPAT 〈<https://www.cnpat.com.cn/Oneshot/index/id/597.html>〉

⁴¹ LIU Yi (2022) 「中国における AI の知的財産への応用」
〈<https://www.japio.or.jp/fair/files/2022/202205.pdf>〉

③対象言語

対象言語は、不明である。

④追加機能等

追加機能については、新語への対応を図っている。

⑤その他

前述のとおり、中国特許情報センター（CNPAT）では、2008年より、中英機械翻訳システムを提供し、2019年からは、NMT翻訳サービスを提供した。

(5) 韓国特許庁（KIPO）の機械翻訳機能の提供例

①サービスの概要

韓国の特許情報システムとして、韓国特許庁（KIPO）の韓国特許情報オンラインネットワーク（K-PON）⁴²と、韓国特許庁（KIPO）、韓国特許情報院（KIPI）が運営する特許情報検索サービス「KIPRIS」では、韓国の特許・実用新案公報を検索することができ、特許公報の機械翻訳することができる^{43,44}。

韓国特許庁の機械翻訳エンジンは、2018年にWIPOと韓国IT企業との業務提携を通じて、技術移転したものを、独自開発した。特許分野に最適化した翻訳エンジンを開発するために、特許公報、審査および審判文書から約1,033万件の学習データを構築し、学習させたものである⁴⁵。

②翻訳エンジン・アルゴリズム・訳質

翻訳エンジンの詳細情報は公開されていない。一方で、韓国特許庁の機械翻訳エンジンの開発にあたっては、WIPOとの業務提携を通じて移転されており、WIPOの機械翻訳サービスは、ニューラル機械翻訳モデルを採用しているため、アルゴリズムは、ニューラルネットワークによる高精度のアルゴリズムを用いているものと推察される。

③対象言語

対象言語は、K-PIONについては、韓国語⇔英語であり、KIPRISは、韓国⇔英語、韓国⇔日・英の言語に対応している⁴⁶。一方で、本調査対象言語には対応していない。

⁴² https://www.kipo.go.kr/en/HtmlApp?c=97003&catmenu=ek07_06_02

⁴³ 新興国等知財情報データベース「韓国における特許検索方法－特許情報検索サービス（KIPRIS）」, 2022年10月20日。〈<https://www.globalipdb.inpit.go.jp/etc/26764/>〉

⁴⁴ KIPRIS Plus 〈<https://plus.kipris.or.kr/eng/main.do>〉

⁴⁵ JETRO（2022）「韓国特許庁、欧州特許公報韓国語翻訳文データ500万件を開放」, 2022年12月22日。〈<https://www.jetro.go.jp/world/asia/kr/ip/ipnews/2022/221222.html>〉

⁴⁶ 特許庁（2009）「特許庁の特許情報提供における機械翻訳の活用と今後の課題」, 〈[https://aamtjapio.com/kenkyu/files/kenkyu03/AAMT_Japio_sympto\(20091127\)-01.pdf](https://aamtjapio.com/kenkyu/files/kenkyu03/AAMT_Japio_sympto(20091127)-01.pdf)〉

④追加機能等

追加機能については、情報は公開されていない。

⑤その他

2022年12月22日に、KIPRIS Plus を通じて、欧州特許公報の韓国語翻訳文データを民間に無料開放した。2023年中に、米国特許公報（1700万件）を開放した。

(6) 世界知的所有権機関（WIPO）の機械翻訳機能の提供例

①サービスの概要

世界知的所有権機関（WIPO）では、WIPO Translate を開発した⁴⁷。WIPO Translate は、WIPOの先端技術応用センター（Advanced Technology Applications Center：ATAC）により開発された機械翻訳ツールであり、AI機能を用いて、特許文書、科学論文、様々な技術コンテンツの自動翻訳を行うことができる。

WIPO Translate の特徴は、膨大な多言語データに基づいてトレーニングされた最先端のニューラル機械翻訳モデルを利用している。また、ドメイン固有の用語にも対応している。専門的な語彙や業界特有の用語を学習しており、特許明細書、特許請求の範囲、技術的内容の翻訳を実現している。WIPO Translate では、文書の機密性を保護し、翻訳プロセス全体を通じてデータの安全性を確保するためのセキュリティ対策も行っている。

WIPO Translate は、PATENTSCOPEをはじめ、WIPOのデータベース・情報提供サービスの利用の際に利用することができ、自由入力文の機械翻訳や特許情報の機械翻訳サービスを提供している。

②翻訳エンジン・アルゴリズム・訳質

翻訳エンジンの詳細情報は公開されていない。アルゴリズムについては、ニューラル機械翻訳モデルを採用していることから、ニューラルネットワークによる高精度アルゴリズムを用いているものと推察される。

③対象言語

対象言語は、アラビア語、中国語、英語、フィンランド語、フランス語、ドイツ語、イタリア語、日本語、韓国語、ポーランド語、ポルトガル語、スペイン語、ロシア語の13言語に対応している。本調査対象言語では、スペイン語、ポルトガル語、イタリア語、アラビア語に対応しているものの、ASEAN地域の言語への対応は行われていない。

④追加機能等

追加機能についての情報は、見られなかった。

⁴⁷ WIPO Translate <<https://www.wipo.int/web/ai-tools-services/wipo-translate>>

⑤その他

WIPO Translate は、2016年9月に、日英/英日および仏英/英仏の長文翻訳サービスの提供が始まり、ドイツ語、英語、スペイン語、フランス語、日本語、韓国語、ロシア語、中国語の書誌データ翻訳の提供を開始した。同年10月31日からは、AIベースの特許文書の翻訳ツールを開発した（WIPO Translate に組み込む）。2017年9月からは、AIベースの特許文書翻訳の対象言語が拡大された（PCT 公用語⇔英語）⁴⁸⁴⁹。

2.3.2. 特許特化型機械翻訳サービス・エンジンの事例

特許特化型機械翻訳サービスについては、みんなの自動翻訳（2.2. 汎用機械翻訳エンジンで記載）、化学情報協会、サイバーパテント、日本特許情報機構、日本特許翻訳株式会社の5つのサービスを調査した。以下、各事例について示す。

（1）みんなの自動翻訳

前述のとおり「みんなの自動翻訳」は、国立研究開発法人情報通信研究機構（NICT）が開発した自動翻訳技術であり、AI 技術で多用される深い階層構造を持つネットワークを用いた自動翻訳技術（AI 自動翻訳）である。「みんなの自動翻訳」は、TexTra として、Word や PowerPoint のファイルを直接翻訳する機能を搭載している。

法人向けの商用版（川村インターナショナル社が提供）では、特許、特許審決等に対応した機械翻訳サービスを提供している。

※翻訳エンジン・アルゴリズム・訳質、対象言語、追加機能等の詳細は「2.2.8. みんなの自動翻訳」を参照のこと。

（2）日本特許情報機構

①サービスの概要

日本特許情報機構（Japio）は、Japio-AI 翻訳を提供し、海外特許出願原稿の翻訳、請求項や明細書の補正案の翻訳、公報の特定箇所（実施例）の翻訳、補正書の翻訳等で活用されている。具体的には、翻訳したいテキスト文書を、特許公報に特化した高精度 AI 機能を活用して高品質な機械翻訳をリアルタイムに提供する。

本サービスでは、各国特許庁が発行する XML 形式の特許文献データを翻訳するために

⁴⁸ WIPO（2016）「WIPO Develops Cutting-Edge Translation Tool For Patent Documents」, 2016年10月31日. 〈https://www.wipo.int/pressroom/en/articles/2016/article_0014.html〉

⁴⁹ WIPO（2017）「WIPO Translate: Cutting-Edge Translation Tool For Patent Documents Extends Language Coverage」, 2017年9月25日. 〈https://www.wipo.int/pressroom/en/articles/2017/article_0007.html〉

Japio が独自開発した機械翻訳フレームワークである。長文の請求項に対する翻訳結果が正確で理解しやすくなるように、自動前編集を行う機能や、特許文献に頻出する上付、下付文字の訳文での高い再現性を有する。

②翻訳エンジン・アルゴリズム・訳質

Japio では、特許公報に特化した AI エンジンを実社開発し提供している。機械翻訳エンジンは、第 2 世代のニューラル機械翻訳技術である Transformer 方式を採用している。これにより、構文認識精度が向上した。また、特許文献に多い長文の翻訳が、従来方式よりも高精度で行えるようになった⁵⁰。

③対象言語

対象言語は、「日本語⇒英語、中国語（簡体字、繁体字）、韓国語」、「英語、中国語（簡体字、繁体字）、韓国語、ドイツ語、フランス語、ロシア語、スペイン語、ポルトガル語⇒日本語」である。本調査の対象言語では、スペイン語、ポルトガル語に対応しているものの、ASEAN、中東地域の言語には対応していない。

表 12 本調査対象言語への対応状況 (Japio)

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | ○ | ○ |
| ポルトガル語 | ○ | ○ |
| イタリア語 | × | × |
| インドネシア語 | × | × |
| マレー語 | × | × |
| ベトナム語 | × | × |
| タイ語 | × | × |
| アラビア語 | × | × |
| ヒンディー語 | × | × |
| トルコ語 | × | × |

④追加機能等

関連する情報は、見られなかった。

⑤その他

本機械翻訳サービスの利用料金については、フリーテキスト翻訳で 100 万文字/月（11000 円/月）（税込み）である。また、オプションサービスとして、公報翻訳サービス（3300 円/月）がある。

また、Japio-GPG/FX と連携し、指定した公報番号の翻訳を、特許文書の翻訳に適するように調整した AI 翻訳機能を使ってリアルタイムに行うサービスがある。

⁵⁰ Japio-AI 翻訳 〈<https://www.japio.or.jp/service/service06.html#006-01>〉

(3) 化学情報協会⁵¹

①サービスの概要

化学情報協会では、化合物名や科学技術専門用語を収載したシソーラス付き対訳辞書、および協会独自の化合物表記翻訳技術を活かした機械翻訳サービスを提供している。

本機械翻訳サービスでは、海外特許・文献の内容把握や、科学技術・医薬製薬分野の翻訳文書作成を効率化するため、2つのサービスを提供している。現行のサービスは、2023年12月に大幅機能強化し、「JAICI AutoTrans」、「JAICI ProTranslator Express Light」を提供している。

「JAICI AutoTrans」は、海外特許・文献の内容把握に最適な機械翻訳サービスであり、アジア言語を含む29言語に対応している。翻訳エンジンは、複数の汎用の翻訳エンジンを活用している。

「JAICI ProTranslator Express Light」は、科学技術・医薬製薬分野の翻訳文書作成を支援・効率化するサービス（機械翻訳を下訳として翻訳文書作成を行う）を提供している。また、機械翻訳に加え、自社で保有する用語集や対訳集といった翻訳資産を活用した専用の翻訳環境の構築や、複数名、企業間でのコラボレーションによる翻訳文書作成も可能である。対象言語は、同サービスの機能別に異なる。

②翻訳エンジン・アルゴリズム・訳質

化学情報協会の機械翻訳エンジンについては、特許出願業務に特化したサービス（DocSpread 翻訳、PatSpread 翻訳）では、NICT 特許特化型機械翻訳エンジンを利用している。一方で、テキスト翻訳等では、汎用機械翻訳エンジンを利用しているため、インドネシア語、ベトナム語、タイ語、ミャンマー語等のアジア言語に対応している。

アルゴリズムについては、特許特化型機械翻訳サービスにおいて NICT のニューラル機械翻訳モデルを採用していることから、ニューラルネットワークによる高精度アルゴリズムを用いているものと推察される。また、汎用機械翻訳エンジンも NMT を採用している。

③対象言語

「JAICI AutoTrans」では、日本語⇄英語、中国語、韓国語等の主要な言語方向へ対応している。また、一部メニューでは、独日・仏日・中英・英中翻訳が可能である。

「JAICI ProTranslator」は、より多くの言語方向に対応している。対応言語は、日本語⇄英語・中国語・韓国語に加え、欧州言語やアジア言語にも対応可能である。ただし、言語の対応は、「JAICI ProTranslator」内の提供サービスにより異なる。翻訳対象文書に合わせ特定の文書を学習させた翻訳エンジンを選択することが可能である。

- [DocSpread 翻訳] NICT 特許特化型エンジンが適用される。翻訳言語は、英語⇄日本

⁵¹ 化学情報協会「機械翻訳サービス」〈<https://www.jaici.or.jp/translation/>〉

語、中国語（簡体字）⇔日本語、中国語（繁体字）→日本語、韓国語⇔日本語、欧州言語⇔日本語である。欧州言語については、フランス語、ドイツ語のみ対応している。

- [テキスト翻訳] 汎用機械翻訳エンジンが適用される。翻訳言語は、英語・中国語・韓国語・欧州言語（フランス語、ドイツ語、スペイン語、ポルトガル語、イタリア語、オランダ語、ポーランド語、ロシア語）・アジア言語（インドネシア語、ベトナム語、タイ語、ミャンマー語）⇔日本語である。
- [PatSpread 翻訳] NICT 特許特化型エンジンが適用される。英語、中国語、韓国語、ドイツ語、フランス語⇔日本語である。
- [エクセル翻訳] NICT 特許特化型エンジンが適用される。英語、中国語、韓国語⇔日本語である。

表 13 本調査対象言語への対応状況（化学情報協会）

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | △汎用 | △汎用 |
| ポルトガル語 | △汎用 | △汎用 |
| イタリア語 | △汎用 | △汎用 |
| インドネシア語 | △汎用 | △汎用 |
| マレー語 | × | × |
| ベトナム語 | △汎用 | △汎用 |
| タイ語 | △汎用 | △汎用 |
| アラビア語 | × | × |
| ヒンディー語 | × | × |
| トルコ語 | × | × |

④追加機能等

化学情報協会の機械翻訳サービスの独自技術として、化合物表記翻訳が挙げられる。一方で、テキスト翻訳では適用対象外である。

⑤その他

2023年12月に、化学情報協会の機械翻訳サービスは、機能強化し、「JAICI ProTranslator Express Light」を提供した。

(4) サイバーパテント⁵²

①サービスの概要

Questel パテントサービスでは、特許の調査、翻訳、国際出願、EP バリデーション、欧州単一効特許の管理、記録、更新等の管理業務を支援している。Orbit Intelligence データベースには、1億件以上の特許、1,700万件以上の意匠、1億5,000万件以上の非特許文献（臨床試験、プロジェクト、科学文献など）が含まれており、検索機能と解析機能で多くの情報

⁵² サイバーパテントデスク 〈<https://www.patent.ne.jp/service/patent/search.html>〉

に基づいた意思決定を支援している。サイバーパテント社は、仏・Questel グループに属する。

CyberPatent Desk では、2022 年 11 月 1 日より、川村インターナショナルが提供する高精度 AI 翻訳システム「みんなの自動翻訳@KI」を適用し、海外出願や海外特許の調査業務での利用できるようにした⁵³。

② 翻訳エンジン・アルゴリズム・訳質

アルゴリズムについては、特許特化型機械翻訳サービスにおいて NICT のニューラル機械翻訳モデルを採用していることから、ニューラルネットワークによる高精度アルゴリズムを用いているものと推察される。

訳質については、「みんなの自動翻訳@KI」の特許エンジンは、特許庁から提供された対訳データで追加学習を行った特許分野全般の翻訳に特化したエンジンであり、機械翻訳の評価に用いられる BLEU スコアでも 50 と高い数値を出している。

③ 対象言語⁵⁴

商用版（特許）の対象言語は、日本語⇄英語、簡体字、繁体字、韓国語、ドイツ語、フランス語のほか、英語⇄簡体字、ドイツ語、フランス語に対応する。

商用版（特許審決等）の対象言語は、日本語⇄英語、簡体字、繁体字に対応する。本調査対象言語には対応していない。

表 14 本調査対象言語への対応状況（化学情報協会）

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | × | × |
| ポルトガル語 | × | × |
| イタリア語 | × | × |
| インドネシア語 | × | × |
| マレー語 | × | × |
| ベトナム語 | × | × |
| タイ語 | × | × |
| アラビア語 | × | × |
| ヒンディー語 | × | × |
| トルコ語 | × | × |

④ 追加機能等

非商用版では、自動翻訳結果の修正、対訳集の登録が可能であり、専用の自動翻訳エン

⁵³ Questel (2022) 「知財情報検索サービス「CyberPatent Desk」高精度 AI 翻訳システムを適用」. (<https://prtimes.jp/main/html/rd/p/000000002.000103736.html>)

⁵⁴ 川村インターナショナル (2023) 「「みんなの自動翻訳@KI (商用版)」に「特許審決等」 「ニュース」の翻訳エンジンを追加」. (<https://prtimes.jp/main/html/rd/p/000000157.000031187.html>)

ジンを登録・訓練することが可能。また、用語集で訳語を指定しカスタムした自動翻訳を構築することができる。

カスタム自動翻訳では、複数の自動翻訳エンジン（汎用、特許、対話、FSA、金融、法令契約、汎用+実験用・技術移転不可）を同時に使うことが可能である。

⑤その他

特許特化型の機械翻訳エンジン（みんなの自動翻訳@KI）は、前述のとおり、2022年11月にサービスが提供された。

サイバーパテントでは、CyberPatent Deskに「みんなの自動翻訳@KI」を適用するにあたり、同社の求める性能を標準のパッケージで実現することは不可能であったが、川村インターナショナルの協力により、「特許エンジン」の翻訳品質はそのままに、必要な可用性とスループットを実現する専用のシステムを短期間に構築することができたとされる。

(5) 日本特許翻訳株式会社⁵⁵

①サービスの概要

日本特許翻訳株式会社では、NICT 新世代 NMT をベースに、機能拡張されたメモソースにより、翻訳業務を支援する「ProTranslator」を提供している⁵⁶。2023年9月には、ProTranslator EXPRESS/EXPRESS-Light は、一般社団法人化学情報協会（JAICI）からのライセンス提供により、化合物表記翻訳を正確に行うことを可能とする化合物表記翻訳に対応した。

②翻訳エンジン・アルゴリズム・訳質

機械翻訳エンジンは、NICT 特許、汎用エンジン+独自エンジン（Domain adaptive MT）を用いている。特に、NICT 特許 NT エンジンをベースエンジンとして、顧客対訳データや翻訳メモリなどから NICT アダプテーション+EBMT エンジン（npat I NICT エンジン）を構築し、Trados/Memsource/MemoQ などの CAT ツールと連携した統合型翻訳環境を提供している（NICT アダプテーション+EBMT（NICT 第3世代エンジン））。また、英日-日英のみのアダプテーションエンジン（npat II Express エンジン）がある。

アルゴリズムについては、汎用 NT 又は特許 NT（API、テキスト翻訳）、汎用 NT+独銜 NT（マルチNMT）、汎用NT+独銜NT+アダプテーション（マルチNMT）は、Transformer型であり、ニューラルネットワークによる高精度アルゴリズムを用いているものと推察される。

訳質については、npat I では、評価値が BLEU（65.3）、RIBES（90.6）、Score（BLEU+RIBES/2）（79.0）であった（翻訳バンクに対訳提供が必要）。npat II では、評価値が BLEU

⁵⁵ 日本特許翻訳株式会社 〈<https://npat.co.jp/>〉

⁵⁶ 日本特許翻訳株式会社（2023）「統合翻訳支援サービス ProTranslator Express」, 2023年6月1日. 〈https://npat.co.jp/shared/img/pdf/ProTranslator_EXPRESS_20230601.pdf〉

(65.8)、RIBES (90.5)、Score (BLEU+RIBES/2) (79.6) であった (翻訳バンクに対訳提供が不要)。NICT (分野特化型：日英特許 NT+MT) では、BLEU (44.6)、RIBES (83.6)、Score (BLEU+RIBES/2) (67.5) であった。

③対象言語

- [npat I] NICT 特許 NT、NICT 汎用 NT、DeepL、NICT サイエンス、NICT 拒絶理由通知・審決に係るものの対象言語は、「NICT 特許 NT」で日本語⇄英語、簡体字、繁体字、韓国語、ドイツ語、フランス語に、「NICT 汎用 NT」では、テキスト翻訳のみであるが、日本語⇄ベトナム語、タイ語、ミャンマー語、インドネシア語に対応している。「DeepL」は、日本語⇄ドイツ語、フランス語、スペイン語、ポルトガル語、イタリア語、ロシア語に対応している。「NICT サイエンス」は日本語⇄英語に留まる。「NICT 拒絶理由通知・審決」は、日本語⇄英語、簡体字である。
- [npat II] EXPRESS エンジン、日本語⇄英語に対応している。

表 15 本調査対象言語への対応状況 (日本特許翻訳)

| | 日本語→当該言語 | 当該言語→日本語 |
|---------|----------|----------|
| スペイン語 | △汎用 | △汎用 |
| ポルトガル語 | △汎用 | △汎用 |
| イタリア語 | △汎用 | △汎用 |
| インドネシア語 | △汎用 | △汎用 |
| マレー語 | × | × |
| ベトナム語 | △汎用 | △汎用 |
| タイ語 | △汎用 | △汎用 |
| アラビア語 | × | × |
| ヒンディー語 | × | × |
| トルコ語 | × | × |

④追加機能等

Platinum 版では、MemsourcesEnterprise 版がバンドルされ、PM 数、Linguist 数無制限の利用が可能となる。さらに、出願人 (複数) と分野を指定して、公開特許ファミリーからアダプテーション用コーパス作成を行う。

⑤その他

日本特許翻訳の機械翻訳サービスは、2017 年 1 月 18 日に開始した。2023 年 2 月には、NICT 特許英日・日英、NICT 汎用英日・日英が新エンジン(GPMT-3.9_230131_nmt) にバージョンアップした。また、2023 年 10 月より、ProTranslator EXPRESS では、翻訳者が行ってきたポストエディット作業を生成 AI (OpenAI 社 API) で代替する機能を実装予定であることを発表した⁵⁷。

⁵⁷ 薬事日報「医療翻訳サービス：日本特許翻訳」, 2023 年 9 月 27 日。
<https://www.yakuji.co.jp/entry105604.html>

料金は、ProTranslator (Express Light) は月額 9,500 円からであり、ProTranslator (Express) は、月額 60,000 円からである。

3. 日英中韓 以外の言語をネイティブとする 知財庁の状況に関する調査

3.1. 調査の概要

3.1.1. 調査の目的

特許審査ハイウェイ（PPH）に限らず、各国知財庁審査官は、特許審査を行うにあたって、同等の内容で他国に出願され、先に他国の審査官によって特許審査が行われた結果を参照することにより、知財庁間でのワークシェアリングを実現している。他国の公報・審査結果を参照する場合、英語以外の言語は、基本的には英語に翻訳されてその内容を理解することが多いと考えられ、英語以外の言語をネイティブとする審査官は、既に業務において英語の読み書きに精通していることが考えられる。

本項目では、日英中韓以外の言語をネイティブとする知財庁の審査官計 10 名程度に対して、ネイティブ言語で公報・審査結果を参照する場合と比較し、英語で他国の公報・審査結果を参照する場合の理解度、問題点（訳質、英語力等）、日本国特許庁が提供する日英機械翻訳サービスへの評価、他庁への要望（以上は必須事項）やその他について、アンケート調査を行った。

3.1.2. 調査項目等

アンケート調査項目は、特許情報サービスにおける日英翻訳機能、公報・審査結果の英語での理解、審査業務と英語力の 3 つからなる。

特許情報サービスにおける日英翻訳機能では、日本の審査経過情報を英語で参照する際に利用しているサービス、審査経過情報の翻訳の訳質の評価、サービスの利用の理由、J-PlatPat を利用していない理由、日本の審査経過情報の参照頻度、翻訳サービスの改善点等を質問した。

公報・審査結果の英語での理解では、原文が英語で記載された公報・審査結果の理解度（欧米由来、日中韓由来）を質問するとともに、審査に支障が生じた事例に関する意見を収集した。

審査業務と英語力では、母国語ではなく、英語を用いた審査実務に関する難しさや審査官に必要とされる英語力の指標を質問した。

3.2. アンケート結果分析

3.2.1. 特許情報サービスにおける日英翻訳機能について

(1) 日本の審査経過情報を英語で参照する際に利用しているサービス

日本の審査経過情報を英語で参照する際に利用しているサービスとして最も回答数が多い

いものは、EPO が提供している Espacenet であった。次いで、WIPO が提供する PATENTSCOPE であり、日本の独立行政法人工業所有権情報・研修館が提供する特許情報プラットフォーム（以下、J-PlatPat）と続く。その他と回答したものの中には、STN が提供する The Derwent World Patents Index Database、USPTO の Global Dossier、Google Patent、EPO の Epoquenet 等が挙げられた。

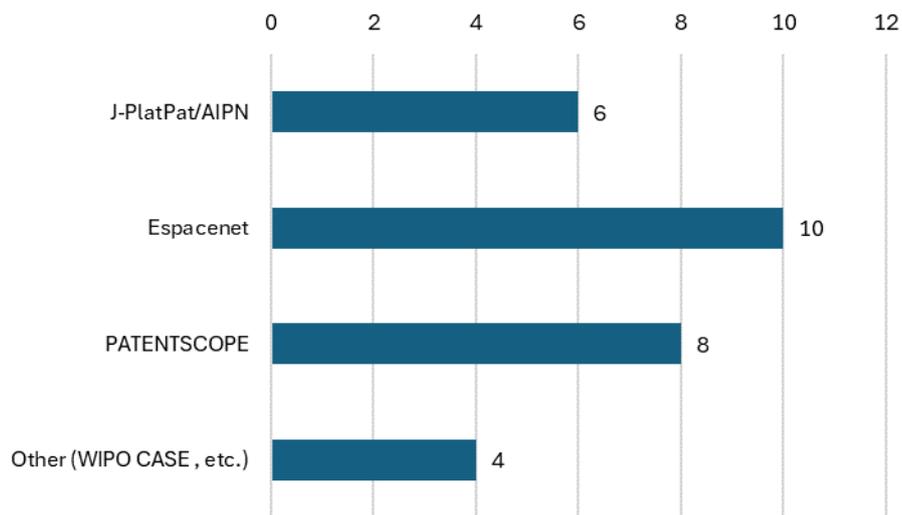


図 4 日本の審査経過情報を英語で参照する際に利用しているサービス（複数回答）

上記で利用しているサービスのうち、日本の審査経過情報を英語で参照する際に最も利用しているものは、EPO が提供している Espacenet であり、次いで、PATENTSCOPE、その他のサービスであった。J-PlatPat を最も利用するとの回答は見られなかった。

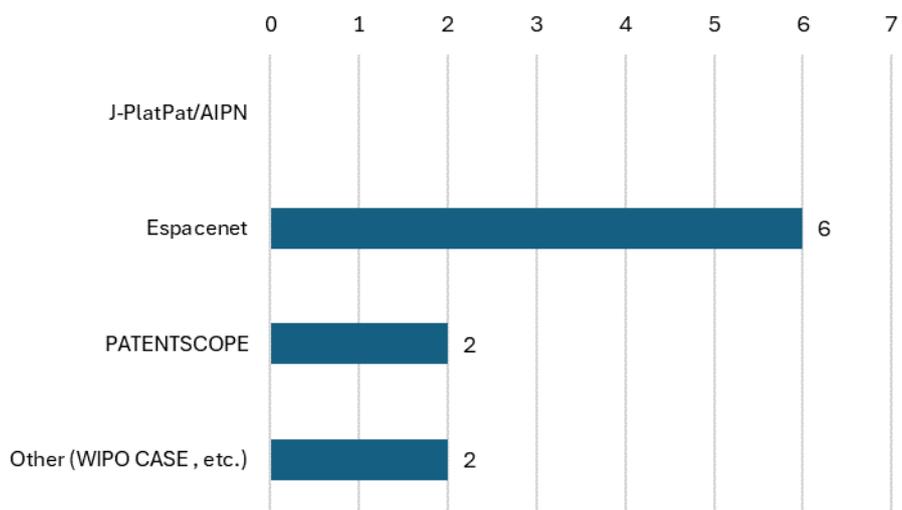


図 5 日本の審査経過情報を英語で参照する際に最も利用しているサービス

(2) 特許情報サービスを利用する際、母国語への翻訳を含め、利用している機械翻訳サービス

前述の特許情報サービスを利用する際に、母国語への翻訳を含め、利用している機械翻訳サービスとして最も利用しているサービスは、Google Translate が最も多い。

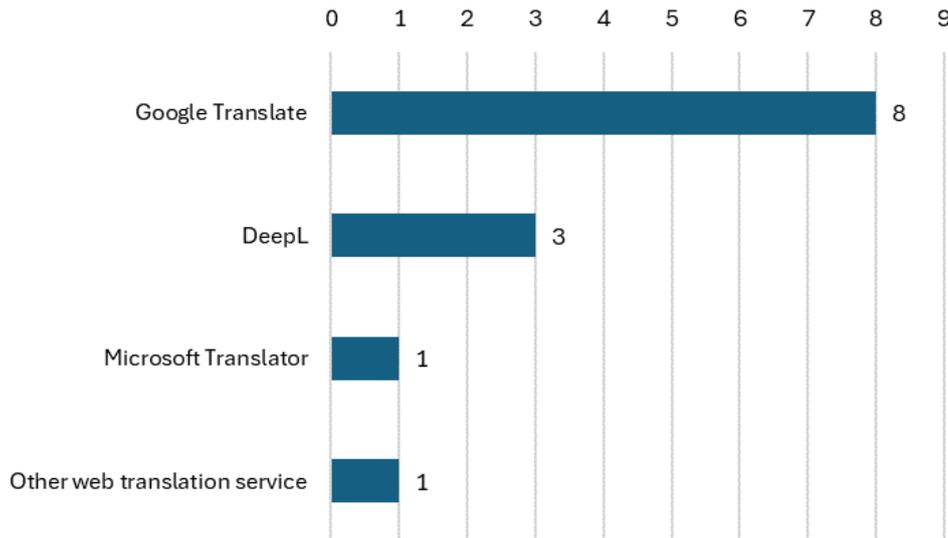


図 6 利用している機械翻訳サービス（複数回答）

日本の審査情報経過を機械翻訳する際の主な利用言語については、英語から母国語に翻訳しているケースが見られる。

表 16 機械翻訳の利用言語（From-To）

| FROM | TO |
|-------------|---------|
| 日本語 | 英語 |
| 日本語 | 英語 |
| 英語 | アラビア語 |
| — | — |
| ドイツ語 | トルコ語 |
| 日本語、韓国語、中国語 | 英語 |
| 英語 | ポルトガル語 |
| 日本語、韓国語、中国語 | 英語 |
| 英語 | ポルトガル語 |
| 英語 | インドネシア語 |

(3) 最も利用しているサービスの審査経過情報の翻訳の訳質の評価（審査への影響等を含む）

前述の日本の審査経過情報を英語で参照する際に最も利用しているサービスの訳質については、「訳質はあまり問題なく審査への支障はあまりない」と認識する回答が最も多い。一方で、「訳質は問題がある場合もあり審査にやや支障がある」と認識している回答も3割見られる。

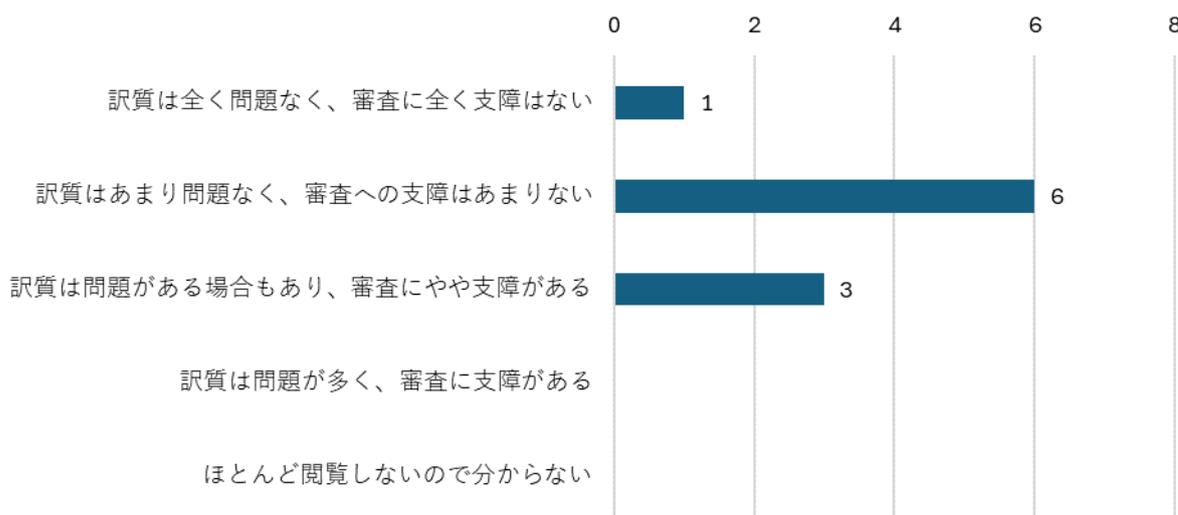


図7 審査経過情報の翻訳の訳質の評価

なお、最も利用しているサービスについての訳質の評価は、最も利用がEspacenetは、審査への支障はあまりないとしつつも、やや支障があるとする回答も2件占める。

表17 最も利用しているサービスについての訳質の評価

| 集計 | 訳質は全く問題なく、審査に全く支障はない | 訳質はあまり問題なく、審査への支障はあまりない | 訳質は問題がある場合もあり、審査にやや支障がある | 訳質は問題が多く、審査に支障がある | ほとんど閲覧しないので分からない |
|-------------------------|----------------------|-------------------------|--------------------------|-------------------|------------------|
| J-PlatPat/AIPN | | | | | |
| Espacenet | 1 | 3 | 2 | | |
| PATENTSCOPE | | 1 | 1 | | |
| Other (WIPO CASE, etc.) | | 2 | | | |
| 総計 | 1 | 6 | 3 | | |

上記の「訳質は問題がある場合もあり審査にやや支障がある」とする回答（3件）のうち、審査に支障が発生した頻度は、審査全体を100とした場合、10、20、70とする回答が1

件ずつであった。

(4) サービスの利用に係る理由（日本の審査経過情報を英語で参照する際の利用のポイント）

サービスの利用者の多い Espacenet の利用理由には、フレンドリーなサービスで、Google 翻訳を併用して利用している意見が複数見られた。

一方で、最も利用しているサービスに「J-PlatPat/AIPN」を挙げる機関は見られなかったが、「J-PlatPat/AIPN」を利用しない理由として、日本発のアプリケーションであること、検索目的に素早くアクセスできない（翻訳表示の遅さを含む）こと、他のサービスと比較して使い勝手が悪い点等が挙げられた。

(5) 日本の審査経過情報の参照頻度

日本の審査経過情報の参照頻度については、「月に複数回利用している」と「数カ月に 1 回程度の頻度で利用している」との回答が同数であり、参照頻度は二極化している。

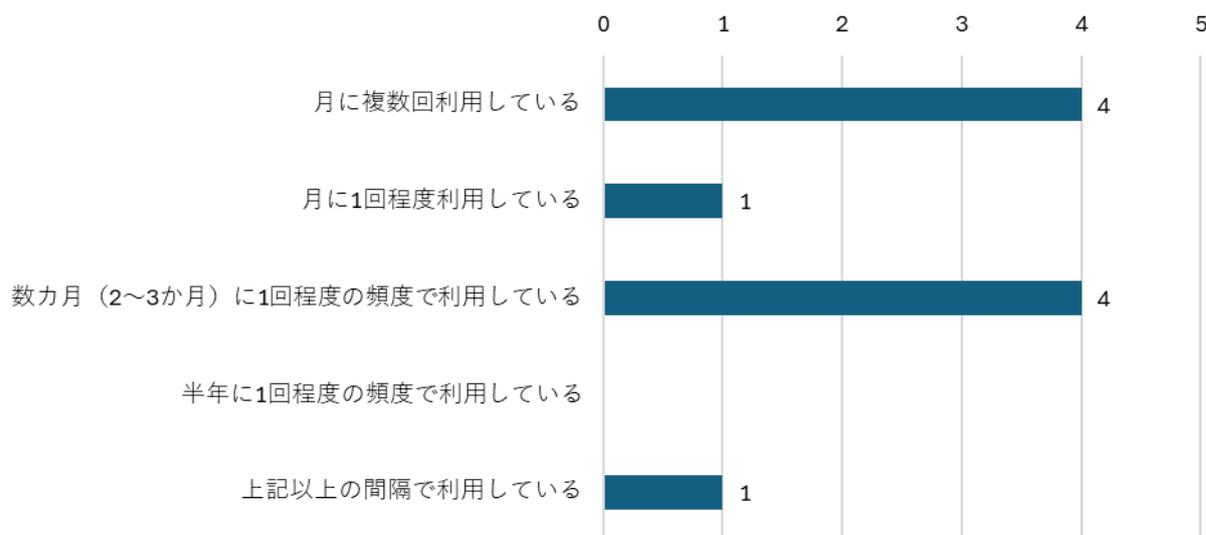


図 8 日本の審査経過情報の参照頻度

(6) 日英翻訳サービス以外で知財庁の翻訳サービスの改善点

日英翻訳サービス以外の知財庁の翻訳サービスの改善点として、言語オプションの追加、英語を介さない翻訳、文の並べ方/単語選択の向上による翻訳品質の向上、翻訳の読みやすさ等の改善点が挙げられた。

3.2.2. 公報・審査結果の英語での理解

①公報・審査結果に対する理解度（審査への影響）

原文が英語で記載された公報・審査結果（例えば、欧米由来）について、ネイティブ言語で公報・審査結果を参照する場合に比べて、公報・審査結果に対する理解度は、「十分理解でき、審査に全く支障がない」とする意見が全体の6割を占めた。

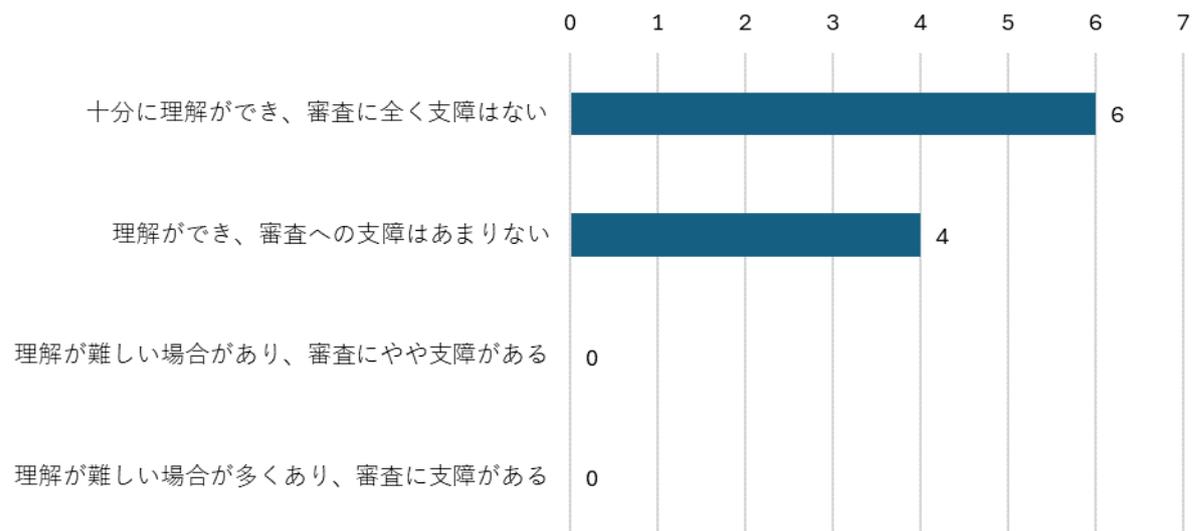


図9 原文（英語）の公報・審査結果の理解度（ネイティブ言語の参照と比して）

原文が英語以外の言語（例えば、日中韓の言語）で英訳した公報・審査結果では、欧米由来の言語と比べて、「十分理解でき、審査に全く支障がない」ではなく「理解でき、審査への支障はあまりない」を選ぶ審査官が増加し、前者を選ぶ審査官は1名のみであった。また、「理解が難しい場合がありやや支障がある」とする回答も2割ほど見られた。

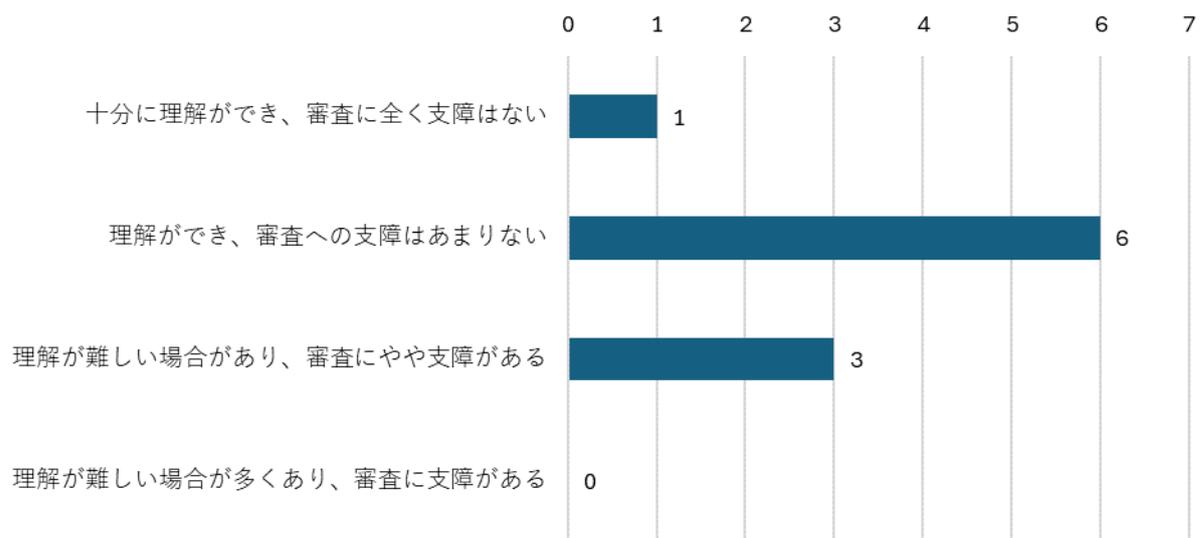


図 10 原文（英語以外：日中韓）の英訳版の公報・審査結果の理解度（ネイティブ言語と比較して）

ネイティブ言語でないため、審査に支障が発生した頻度は、全体を 100 とした場合、2 つの知財庁からそれぞれ支障度 40、50 であると示された。審査に支障が発生した具体的な事例として、翻訳不可能な単語、訳文の文法的な誤り、過去の技術文書の入手が難しい点等が挙げられた。

原文が日本語の翻訳された公報・審査結果については、前述の原文が英語以外の言語（例えば、日中韓の言語）と同じとする回答であった。日本語と他の英語以外の言語の違いについての回答は見られなかった。

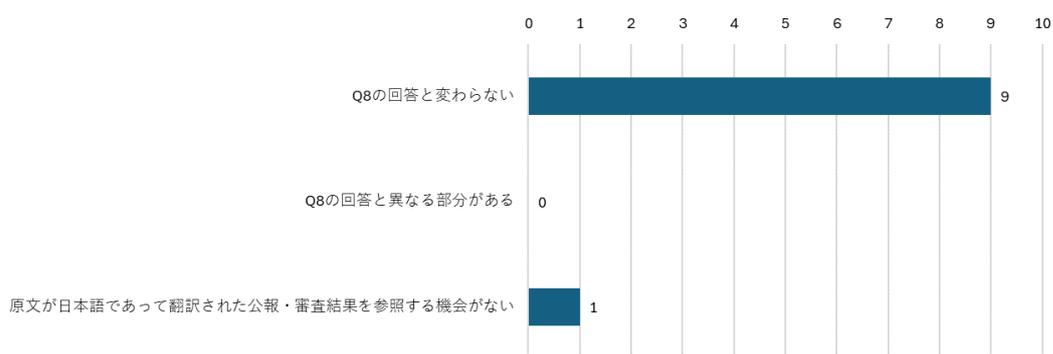


図 11 原文（日本語）の翻訳の理解

注：図中の Q8 は原文（英語以外：日中韓）の英訳版の公報・審査結果の理解度に関する問い（図 10）を指す。

3.2.3. 審査業務と英語力

①母国語ではない言語（英語）を用いた審査実務の課題等

母国語ではなく英語を用いて審査実務（外国庁の公報・審査結果の英訳文の理解、英語の先行技術文献の理解、英語起案等）を行う際に、英語力の点から困難・問題を感じることもあるかについて質問した。ほとんどの知財庁では、審査業務において、英語を用いることについて「困難・問題を感じるということはあまりない」という回答であった。



図 12 母国語でない言語（英語）を用いた審査実務における課題等

②審査官に必要とされる英語力

審査官に必要とされる英語力について、英語力に対する具体的な指標や想定があるとの回答が 6 件、具体的な指標はないとする回答が 2 件であった。

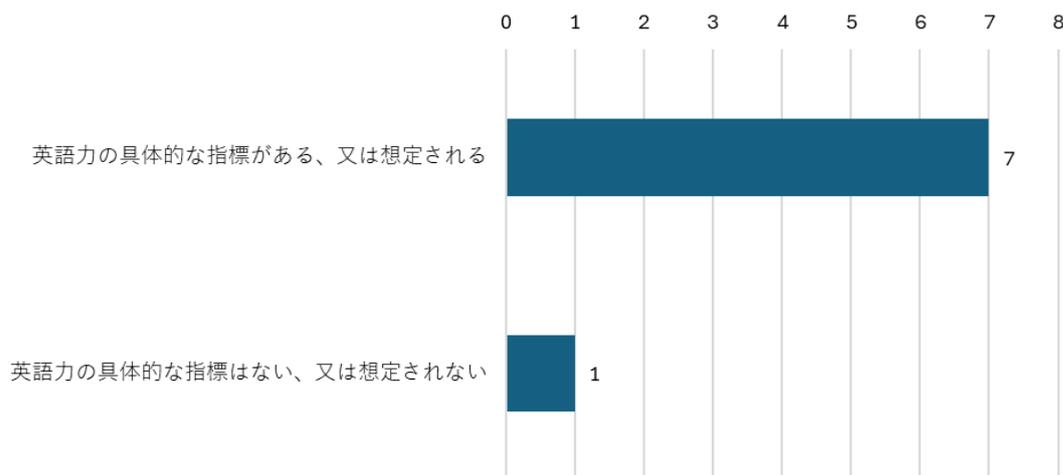


図 13 審査官に必要とされる英語力

英語力について具体的な指標があるとする知財庁の回答を見ると、大学卒業レベル（一

部の国は大学英語テストを挙げている)、TOEIC600 点以上とする回答も見られた。それ以外では、技術文献の英語の理解も重要との意見が挙げられた。

3.3. アンケート結果からの示唆

調査項目 2 では、日英中韓以外の言語を知財庁の審査官を対象にアンケート調査を実施した。

海外の審査官は、日本の審査経過情報を英語で利用する場合、EPO が提供している Espacenet、次いで PATENTSCOPE や J-PlatPat を利用している。また、汎用的な機械翻訳サービスとしては、Google の翻訳サービスを利用している審査官が多い。

調査対象の知財庁の半数以上で、審査官の任用にあたっては、一定の英語力の保持を示す指標があるとされ、多くの審査官は、欧米の公報・審査経過情報を英語で確認することに対しては支障を抱いていないという結果であったが、一方で、原文が英語以外の言語（例えば、日中韓の言語）で英訳した公報・審査結果を確認することには支障の声が一定数あがった。

4. 調査対象言語を使用する国々に関する経済的情報に関する調査

4.1. 調査の概要

本項目では、調査対象言語が知財庁での主要言語になっている国々について、基礎的経済データ（経済規模、日本との間の対外投資・技術貿易収支等）、日本企業の進出状況、知財に係る日本からの出願件数（PPH 申請件数を含む）やその他の経済的情報の収集を行った。

本項目では、調査対象国の経済的情報から、機械翻訳エンジンを活用した特許情報の検索・調査ニーズに関連する、諸外国の経済規模や経済活動状況に係る定量的、定性的情報を収集し、特許情報の機械翻訳の整備に向けた基礎情報を収集した。

(1) 調査対象国

調査対象国は、1.2.2.で示したとおり、スペイン、メキシコ（スペイン語）、ポルトガル、ブラジル（ポルトガル語）、イタリア（イタリア語）、インドネシア（インドネシア語）、マレーシア（マレーシア語）、ベトナム（ベトナム語）、タイ（タイ語）、サウジアラビア（アラビア語）、インド（ヒンズー語）、トルコ（トルコ語）の 10 言語 12 カ国を対象とした。

表 18 調査対象国の基礎情報

| 国名／言語 | 区分 | 内容 |
|------------------|-----|---|
| スペイン *スペイン語 | 面積 | 50 万 5,983 km ² （日本の約 1.3 倍） |
| | 人口 | 4,743 万人（2022 年）※出所:スペイン国立統計局 |
| | 宗教 | カトリック |
| | 公用語 | スペイン語、他に地域ごとの公用語有（カタルーニャ語、ガリシア語、バスク語） |
| | 政体 | 議会君主制 |
| メキシコ *スペイン語 | 面積 | 196 万 4,375 km ² （日本の約 5.2 倍） |
| | 人口 | 1 億 2,601 万人（2020 年国勢調査） ※出所:国立統計地理情報院 |
| | 宗教 | カトリック |
| | 公用語 | 規定なし（事実上、スペイン語） |
| | 政体 | 連邦共和制 |
| ポルトガル *ポルトガル語 | 面積 | 9 万 2,225 km ² （日本の約 4 分の 1） |
| | 人口 | 1,046 万 7,366 人（2022 年） ※出所:ポルトガル国立統計局 |
| | 宗教 | カトリック |
| | 公用語 | ポルトガル語 |
| ポルトガル *ポルトガル語 | 政体 | 共和制 |
| ブラジル *ポルトガル語 | 面積 | 851 万 418 km ² （日本の約 22.5 倍） |
| | 人口 | 2 億 306 万人（2022 年国勢調査） ※出所:ブラジル地理統計院 |
| | 宗教 | カトリック |

| 国名／言語 | 区分 | 内容 |
|--------------------|-----|---|
| | 公用語 | ポルトガル語 |
| | 政体 | 連邦共和制（大統領制） |
| イタリア *イタリア語 | 面積 | 30万2,068 km ² （日本の約5分の4） |
| | 人口 | 5,885万人（2023年） ※出所:イタリア統計局 |
| | 宗教 | キリスト教（ローマ・カトリック、国教ではない） |
| | 公用語 | イタリア語、他に地域ごとの公用語有（ドイツ語、ラディン語、スロヴェニア語、フランス語） |
| | 政体 | 共和制 |
| インドネシア *インドネシア語 | 面積 | 191万6,907 km ² （2019年、日本の約5倍） |
| | 人口 | 2億7,870万人（2023年） ※出所:中央統計庁 |
| | 宗教 | イスラム教、ヒンドゥー教、キリスト教ほか |
| | 公用語 | インドネシア語 |
| | 政体 | 共和制（大統領責任内閣） |
| マレーシア *マレーシア語 | 面積 | 33万411 km ² （日本の0.87倍） |
| | 人口 | 3,301万人（2022年） ※出所:マレーシア統計局 |
| | 宗教 | イスラム教、仏教、キリスト教、ヒンドゥー教等 |
| | 公用語 | マレー語、英語、中国語、タミール語 |
| | 政体 | 立憲君主制 |
| ベトナム *ベトナム語 | 面積 | 33万1,346 km ² （日本の0.88倍） |
| | 人口 | 9,946万人（2022年） ※出所:ベトナム統計総局 |
| | 宗教 | 仏教、その他にカトリック、カオダイ教など |
| | 公用語 | ベトナム語 |
| | 政体 | 社会主義共和国 |
| タイ *タイ語 | 面積 | 51万3,115 km ² （日本の約1.4倍） |
| | 人口 | 6,609万人（2022年） ※出所:内務省 |
| タイ *タイ語 | 宗教 | 人口の約95%が上座部仏教、その他イスラム教（4%）、キリスト教（0.6%）など |
| | 公用語 | タイ |
| | 政体 | 立憲君主制 |
| サウジアラビア *アラビア語 | 面積 | 214万9,700 km ² （日本の約5.7倍） |
| | 人口 | 3,218万人（2022年） ※出所:サウジアラビア統計局 |
| | 宗教 | イスラム教 |
| | 公用語 | アラビア語 |
| | 政体 | 君主制 |
| インド *ヒンディー語 | 面積 | 328万7,263 km ² （日本の約8.8倍） |
| | 人口 | 14億3,651万人（2023年、出所:IMF） |
| | 宗教 | ヒンドゥー教、イスラム教、キリスト教、シーク教徒、仏教など |
| | 公用語 | ヒンディー語（連邦公用語）、英語（準公用語） |
| | 政体 | 連邦共和制 |
| トルコ *トルコ語 | 面積 | 78万43 km ² （日本の約2倍） |
| | 人口 | 8,527万9,553人（2022年） ※出所:トルコ統計機構 |

| 国名／言語 | 区分 | 内容 |
|-------|-----|-------------|
| | 宗教 | イスラム教、キリスト教 |
| | 公用語 | トルコ語 |
| | 政体 | 共和制 |

(2) 経済的情報に関する収集データ

本項目では、調査対象国の経年の経済的状況を把握するために GDP や輸出入額等の基礎的経済データの収集と対象国の経済環境（産業動向等）を把握するためのその他のデータの収集を行った。経年データについては、直近 15 年間（2008～2022 年）までの情報を対象とした。

収集した経済的情報の収集データについては、下記のとおりである。

表 19 経済的情報に関する収集データ

| 種別 | 大項目 | 中項目 | 取得データ |
|-----------------|--------------------|-----------------|-------------|
| 基礎的データ | GDP | 名目 GDP 総額 | 2008～2022 年 |
| | | 一人当たり名目 GDP | 2008～2022 年 |
| | | 購買力平価換算 GDP | 2008～2022 年 |
| | 成長率 | 実質 GDP 成長率 | 2008～2022 年 |
| | 輸出入額 | 対日輸出入額 | 2008～2022 年 |
| | 海外直接投資 | 対外直接投資：日本からの投資額 | 2008～2022 年 |
| | | 対内直接投資：日本への投資額 | 2008～2022 年 |
| | 技術貿易収支 | 日本からの技術輸出 | |
| | | 日本への技術輸入 | |
| | 特許 | 日本からの特許出願件数 | |
| 対象国から日本への特許出願件数 | | | |
| PPH 申請件数 | | 2011 年以降 | |
| その他のデータ | 日本企業の進出状況、ビジネス展開状況 | | *定性的情報 |
| | 日本との関係 | | *定性的情報 |
| | 当該国の製造業の状況 | | *定性的情報 |
| | 地理的・地政学的重要性 | | *定性的情報 |
| | その他のリスク | | *定性的情報 |

基礎的経済データの収集期間については、近年の世界規模の経済的イベント（リーマンショック〈2008年〉、COVID-19〈2020年〉）の影響等を把握するため、直近15年間とした。なお、調査対象期間前後で生じた経済、特許関連の事象は、下記のとおりである。

表 20 調査対象期間、及びその前後の経済、特許関連の注目事象

| 項目 | 内容 |
|----------------|--|
| GDP 関連の各国動向の変化 | <ul style="list-style-type: none"> 2000 年：中国の購買力平価 GDP（PPP 基準、ドル換算）が日本を超えた 2010 年：中国の名目 GDP（ドル換算）が日本を超えた 2018 年：中国の購買力平価 GDP が日本の 4 倍に達した |

| 項目 | 内容 |
|-------------------------|--|
| 一人当たり GDP 関連の変化 | <ul style="list-style-type: none"> 2009年：台湾の1人当たり GDP (PPP 基準) が日本を超えた 2018年：韓国の1人当たり GDP (PPP 基準) が日本を超えた |
| 学術論文関連の変化 ⁵⁸ | <ul style="list-style-type: none"> 2000年代中頃：中国「被引用数上位10%論文数」が日本を超えた 2020年頃から：人口が少ない韓国の「被引用数上位10%論文数」が日本と拮抗 |
| 特許出願関連の変化 | <ul style="list-style-type: none"> 2006年：米国の出願件数が日本の出願件数を再び超えた 2010年：中国の出願件数が日本の出願件数を超えた |

4.2. 基礎的経済データについて

基礎的経済データについては、調査対象国の経済規模を把握するため、GDP（名目 GDP 総額、一人当たり名目 GDP、購買力平価換算 GDP）、経済成長率（実質 GDP 成長率）、輸出入額のデータを収集した。また、日本との間の経済活動を把握するため、対外・対内直接投資を、技術貿易を把握するため、技術貿易収支のデータを収集した。

特許関連のデータについては、日本からの特許出願件数、対象国から日本への特許出願件数、PPH 申請件数を把握した。

4.2.1. 調査対象国の経済規模（GDP、成長率、輸出入額等）

(1) 調査対象国の国内総生産（GDP）

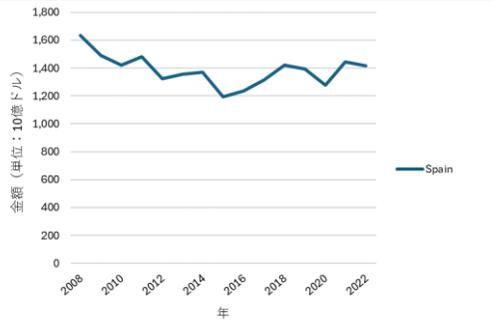
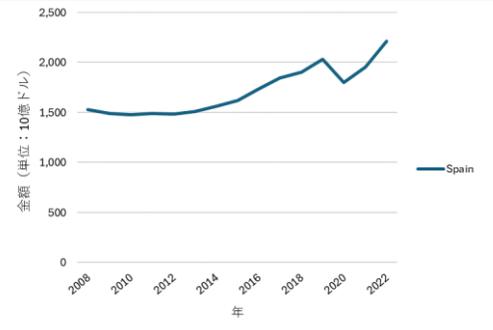
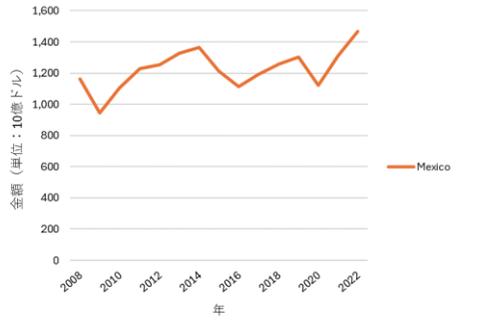
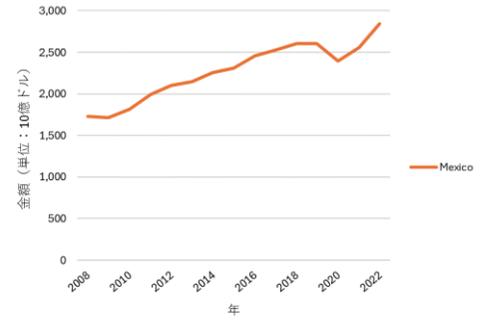
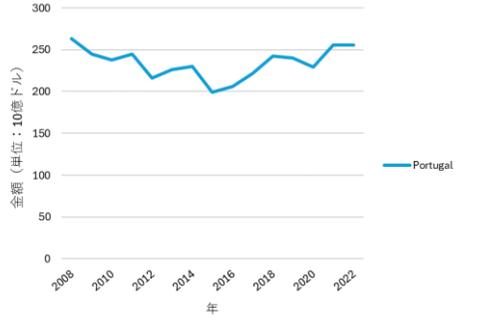
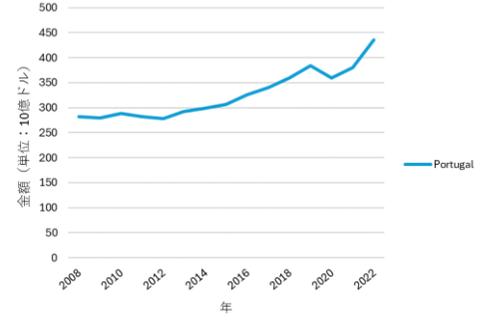
調査対象国の国内総生産（名目、購買力平価換算⁵⁹）は、表 21 のとおりである。対象国別に見ると、名目 GDP、購買力平価換算の GDP の両方とも右肩上がりの国は、インドネシア、マレーシア、ベトナム、タイ等の ASEAN 諸国やインド等が該当する。

⁵⁸ 科学技術・学術政策研究所「4.1.2 研究活動の国別比較（2）国・地域別論文数、Top10% 補正論文数、Top1%補正論文数の時系列比較」

〈https://www.nistep.go.jp/sti_indicator/2022/RM318_42.html〉

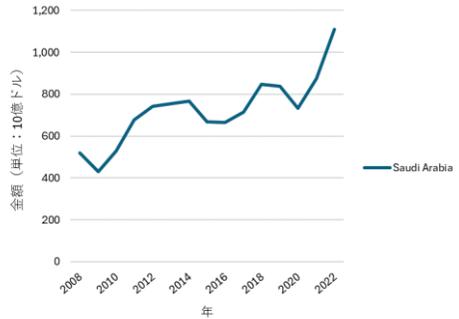
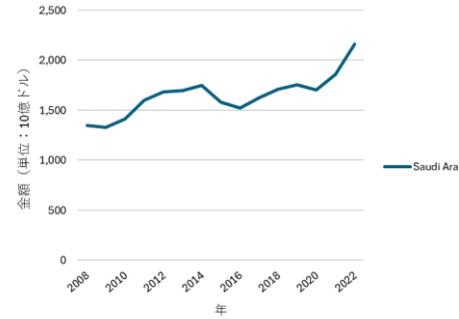
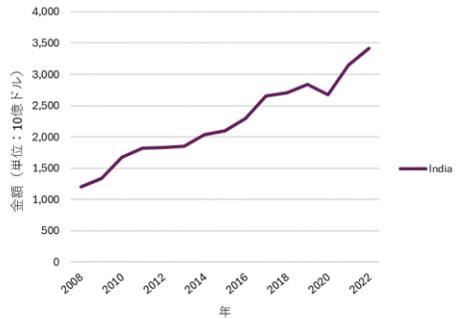
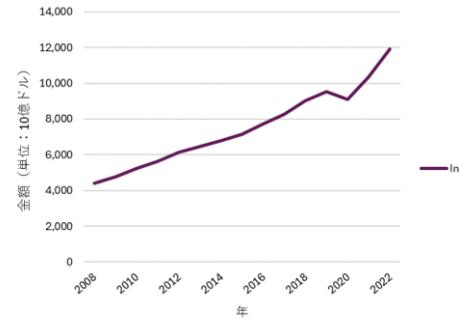
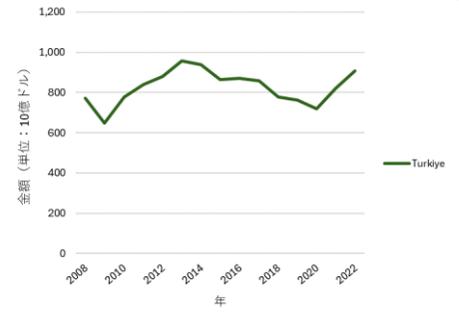
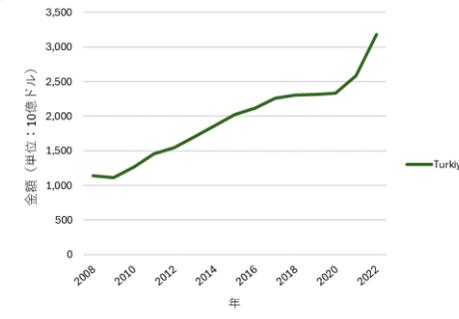
⁵⁹ 購買力平価（PPP）換算は、国際比較を行う際に国内の広い範囲の商品・サービスを反映し、かつ資本異動の影響をうけにくく安定性のある換算レート。

表 21 調査対象国の国内総生産（GDP）の推移 [2008→2015→2022年]（単位：10億米ドル）

| | 名目 GDP 総額 (10億米ドル) 推移 | | GDP (購買力平価換算 PPP) (10億米ドル) | |
|-------|---|--|---|---|
| スペイン | 1,419 [2022年] 1,196 [2015年] 1,632 [2008年] |  | 2,272 [2022年] 1,622 [2015年] 1,499 [2008年] |  |
| メキシコ | 1,466 [2022年] 1,213 [2015年] 1,162 [2008年] |  | 3,064 [2022年] 2,310 [2015年] 1,881 [2008年] |  |
| ポルトガル | 252 [2022年] 199 [2015年] 263 [2008年] |  | 439 [2022年] 307 [2015年] 281 [2008年] |  |

| | 名目 GDP 総額 (10 億米ドル) 推移 | | GDP (購買力平価換算 PPP) (10 億米ドル) | |
|--------|--|--|--|--|
| ブラジル | 1,920 [2022 年] 1,800 [2015 年] 1,696 [2008 年] | | 3,837 [2022 年] 3,015 [2015 年] 2,559 [2008 年] | |
| イタリア | 2,012 [2022 年] 1,837 [2015 年] 2,408 [2008 年] | | 3,059 [2022 年] 2,242 [2015 年] 2,154 [2008 年] | |
| インドネシア | 1,319 [2022 年] 861 [2015 年] 559 [2008 年] | | 4,037 [2022 年] 2,648 [2015 年] 1,813 [2008 年] | |

| | 名目 GDP 総額 (10 億米ドル) 推移 | | GDP (購買力平価換算 PPP) (10 億米ドル) | |
|-------|--|--|--|--|
| マレーシア | 407 [2022 年] 301 [2015 年] 242 [2008 年] | <p>金額 (単位: 10億ドル)</p> <p>年</p> <p>Malaysia</p> | 1,137 [2022 年] 751 [2015 年] 544 [2008 年] | <p>金額 (単位: 10億ドル)</p> <p>年</p> <p>Malaysia</p> |
| ベトナム | 406 [2022 年] 237 [2015 年] 125 [2008 年] | <p>金額 (単位: 10億ドル)</p> <p>年</p> <p>Viet Nam</p> | 1,321 [2022 年] 700 [2015 年] 412 [2008 年] | <p>金額 (単位: 10億ドル)</p> <p>年</p> <p>Viet Nam</p> |
| タイ | 495 [2022 年] 401 [2015 年] 291 [2008 年] | <p>金額 (単位: 10億ドル)</p> <p>年</p> <p>Thailand</p> | 1,482 [2022 年] 1,087 [2015 年] 815 [2008 年] | <p>金額 (単位: 10億ドル)</p> <p>年</p> <p>Thailand</p> |

| | 名目 GDP 総額 (10 億米ドル) 推移 | | GDP (購買力平価換算 PPP) (10 億米ドル) | |
|---------|--|---|---|--|
| サウジアラビア | 1,108 [2022 年] 669 [2015 年] 520 [2008 年] |  | 2,150 [2022 年] 1,578 [2015 年] 1,347 [2008 年] |  |
| インド | 3,390 [2022 年] 2,104 [2015 年] 1,224 [2008 年] |  | 11,901 [2022 年] 7,160 [2015 年] 4,237 [2008 年] |  |
| トルコ | 906 [2022 年] 864 [2015 年] 771 [2008 年] |  | 3,353 [2022 年] 2,023 [2015 年] 1,219 [2008 年] |  |

出所：IMF World Economic Outlook database: October 2023、World Bank（名目 GDP 総額グラフ、GDP（購買力平価 PPP）グラフ）から作成。

(2) 輸出入額

調査対象国の輸出入額については、下記のとおりである。

スペイン語圏、ポルトガル語圏の近年の輸出入額については、スペインは輸出超過（貿易黒字）であるが、ブラジルは均衡しており、メキシコ、ポルトガルは輸入超過（貿易赤字）となっている。各国とも、2020年以降は輸出額、輸入額とも輸出入額を伸ばしている。

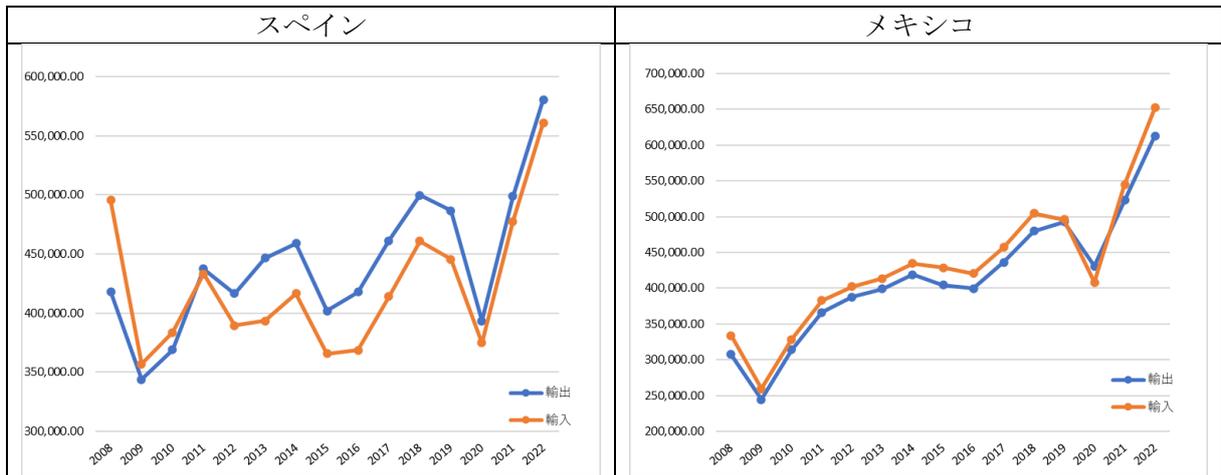


図 14 スペイン語圏の調査対象国の輸出入額（単位：百万ドル）

出所：World Bank national accounts data, and OECD National Accounts data files より作成。

〈<https://data.worldbank.org/indicator/NE.EXP.GNFS.CD>〉

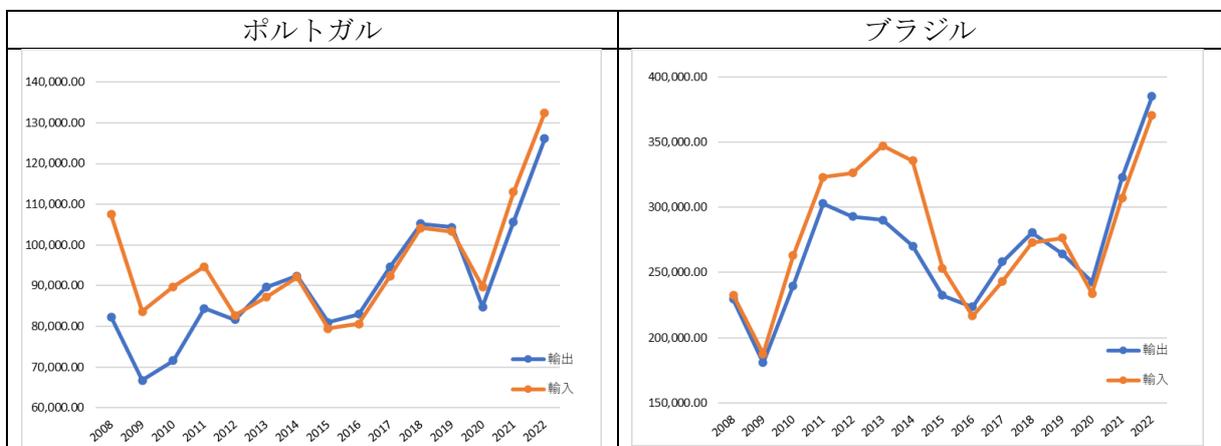


図 15 ポルトガル語圏の調査対象国の輸出入額（単位：百万ドル）

出所：World Bank national accounts data, and OECD National Accounts data files より作成。

〈<https://data.worldbank.org/indicator/NE.EXP.GNFS.CD>〉

イタリアについては、2011年以降、輸出超過の状況であったが、2022年に輸入超過に転じた。

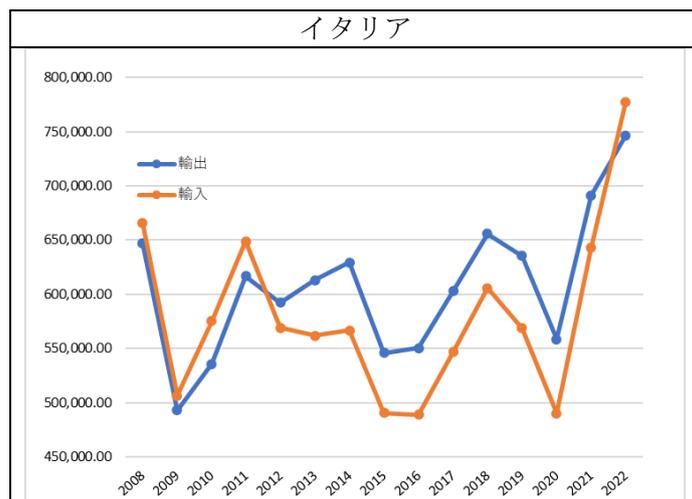


図 16 調査対象国（イタリア）の輸出入額（単位：百万ドル）

出所：World Bank national accounts data, and OECD National Accounts data files より作成。

〈<https://data.worldbank.org/indicator/NE.EXP.GNFS.CD>〉

ASEAN 地域の調査対象国については、インドネシアは 2020 年以降、マレーシアは調査対象年の 2008 年より輸出超過の状況が続いている。

一方で、ベトナムについては、輸出入額は均衡して発展してきている。また、タイについては、直近の 2022 年に輸入超過となった。

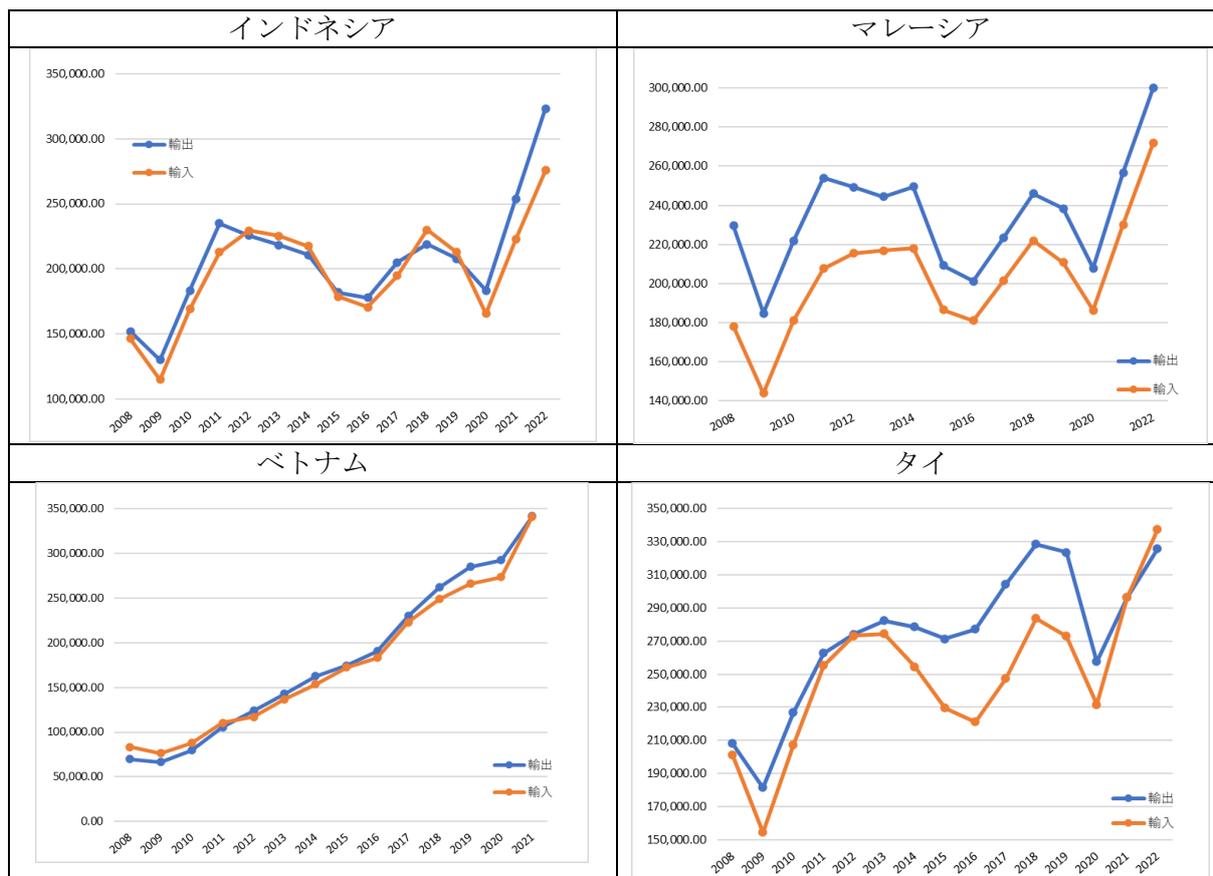


図 17 ASEAN 地域の調査対象国の輸出入額 (単位: 百万ドル)

出所: World Bank national accounts data, and OECD National Accounts data files より作成。

〈<https://data.worldbank.org/indicator/NE.EXP.GNFS.CD>〉

南アジア、中東地域の調査対象国については、インド、トルコは、近年、輸入超過の状況にある。他方、資源国であるサウジアラビアは、輸出超過の状況で推移しているが、COVID-19 の真っ只中の 2020 年は、輸出入額はほぼ同額となった。



図 18 南アジア・中東地域の調査対象国の輸出入額 (単位: 百万ドル)

出所: World Bank national accounts data, and OECD National Accounts data files より作成。

〈<https://data.worldbank.org/indicator/NE.EXP.GNFS.CD>〉

4.2.2. 海外直接投資および技術貿易収支

(1) 対外・対内直接投資

国外で事業活動を行うために企業の買収や、生産設備等への投資に係る指標として、海外直接投資がある。調査対象国の対外・対内直接投資(残高)を見ると、日本から対象国の対外直接投資残高について、最も多い国(2022年)はタイ(747億ドル)であり、次いでインドネシア(357億ドル)、インド(295億ドル)と続く(図19)。

一方で、調査対象国から日本への対内直接投資残高の多い国について、最も多い国はタイ(285億ドル)であり、次いでイタリア(126億ドル)、マレーシア(126億ドル)と続く(図20)。なお、本調査では、2008年からの経年データを収集しているが、対外・対内直接投資額については、国際収支関連統計の基準変更により、2013年以前と2014年以降のデータの連続性はない点に留意する必要がある。

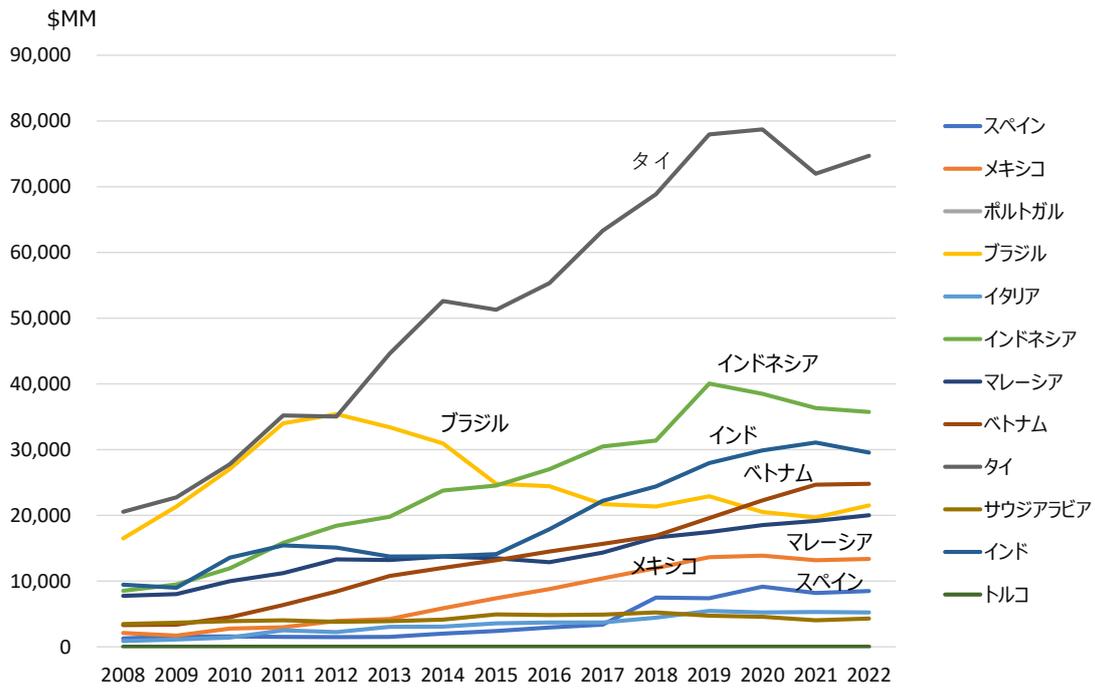


図 19 対外直接投資残高の推移（日本→対象国）

出所：JETRO 直接投資統計（<https://www.jetro.go.jp/world/japan/stats/fdi.html>）より作成。

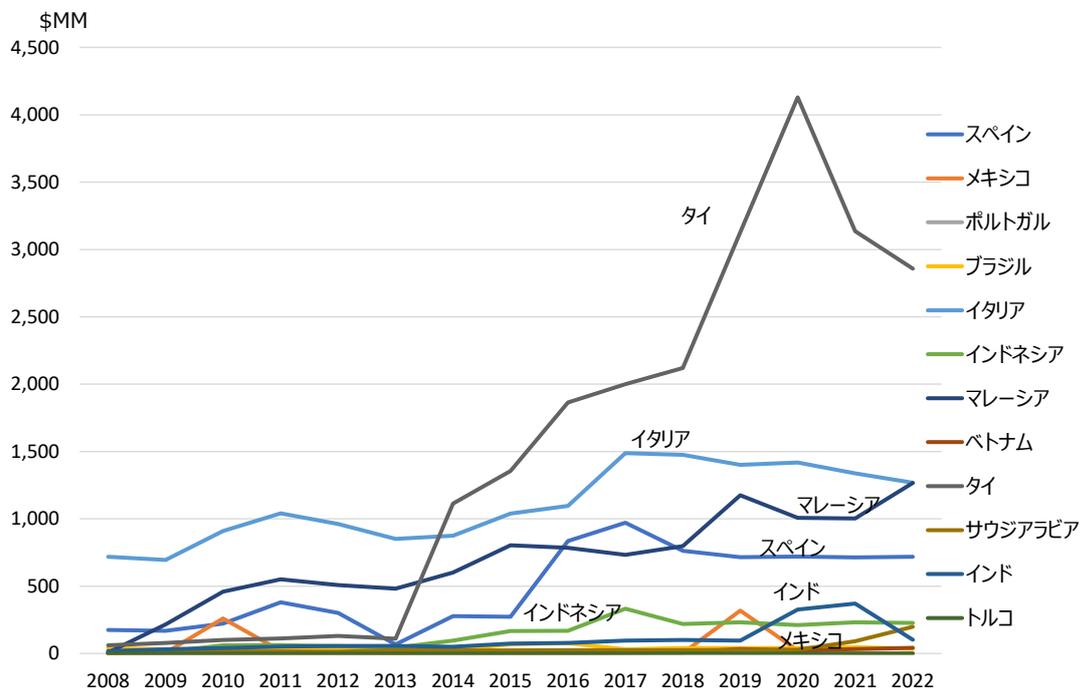


図 20 対内直接投資残高の推移（対象国→日本）

出所：JETRO 直接投資統計（<https://www.jetro.go.jp/world/japan/stats/fdi.html>）より作成。

表 22 対外・対内直接投資（残高）

| 項 目 | 年 | スペイン | メキシコ | ポルトガル | ブラジル | イタリア | インドネシア | マレーシア | ベトナム | タイ | サウジアラビア | インド | トルコ |
|--------------------------------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|--------|------|
| 日本からの各年末の 投資残高 (単位：\$MM) | 2008 | 1,276 | 2,097 | n.a. | 16,492 | 882 | 8,528 | 7,743 | 3,307 | 20,529 | 3,481 | 9,440 | n.a. |
| | 2009 | 1,568 | 1,718 | n.a. | 21,337 | 1,100 | 9,491 | 8,017 | 3,353 | 22,748 | 3,650 | 8,982 | n.a. |
| | 2010 | 1,562 | 2,755 | n.a. | 27,038 | 1,410 | 11,946 | 9,972 | 4,501 | 27,789 | 3,905 | 13,558 | n.a. |
| | 2011 | 1,517 | 2,983 | n.a. | 33,982 | 2,508 | 15,816 | 11,211 | 6,370 | 35,178 | 4,016 | 15,416 | n.a. |
| | 2012 | 1,476 | 3,948 | n.a. | 35,402 | 2,253 | 18,427 | 13,312 | 8,415 | 35,040 | 3,809 | 15,107 | n.a. |
| | 2013 | 1,512 | 4,241 | n.a. | 33,402 | 3,041 | 19,787 | 13,204 | 10,790 | 44,581 | 3,906 | 13,738 | n.a. |
| | 2014 | 2,002 | 5,854 | n.a. | 30,955 | 3,056 | 23,770 | 13,741 | 12,009 | 52,588 | 4,141 | 13,751 | n.a. |
| | 2015 | 2,401 | 7,375 | n.a. | 24,780 | 3,577 | 24,532 | 13,463 | 13,156 | 51,272 | 4,925 | 14,101 | n.a. |
| | 2016 | 2,945 | 8,771 | n.a. | 24,446 | 3,698 | 27,009 | 12,861 | 14,481 | 55,326 | 4,838 | 17,889 | n.a. |
| | 2017 | 3,360 | 10,401 | n.a. | 21,707 | 3,686 | 30,460 | 14,315 | 15,662 | 63,296 | 4,908 | 22,202 | n.a. |
| | 2018 | 7,489 | 11,960 | n.a. | 21,348 | 4,436 | 31,358 | 16,635 | 16,901 | 68,856 | 5,223 | 24,390 | n.a. |
| | 2019 | 7,391 | 13,623 | n.a. | 22,897 | 5,459 | 40,048 | 17,450 | 19,581 | 77,937 | 4,724 | 27,953 | n.a. |
| | 2020 | 9,142 | 13,853 | n.a. | 20,508 | 5,214 | 38,483 | 18,532 | 22,258 | 78,716 | 4,560 | 29,889 | n.a. |
| 2021 | 8,197 | 13,160 | n.a. | 19,688 | 5,302 | 36,332 | 19,147 | 24,679 | 71,986 | 4,035 | 31,072 | n.a. | |
| 2022 | 8,480 | 13,367 | n.a. | 21,498 | 5,215 | 35,731 | 19,998 | 24,797 | 74,702 | 4,309 | 29,549 | n.a. | |
| 日本への各年末の 投資残高 (単位：\$MM) | 2008 | 175 | 6 | n.a. | 40 | 719 | 12 | 7 | 0 | 61 | 4 | 18 | n.a. |
| | 2009 | 168 | 6 | n.a. | 32 | 694 | 12 | 216 | 0 | 79 | 13 | 32 | n.a. |
| | 2010 | 223 | 261 | n.a. | 38 | 909 | 60 | 460 | 0 | 100 | 15 | 40 | n.a. |
| | 2011 | 380 | 20 | n.a. | 41 | 1,041 | 62 | 551 | 0 | 111 | 16 | 52 | n.a. |
| | 2012 | 302 | 18 | n.a. | 35 | 962 | 55 | 509 | 0 | 130 | 14 | 56 | n.a. |
| | 2013 | 68 | 5 | n.a. | 51 | 851 | 44 | 481 | 0 | 110 | 25 | 56 | n.a. |
| | 2014 | 277 | 17 | n.a. | 25 | 875 | 95 | 602 | 1 | 1,114 | 22 | 49 | n.a. |
| | 2015 | 272 | 15 | n.a. | 69 | 1,039 | 167 | 804 | 1 | 1,355 | 23 | 74 | n.a. |
| | 2016 | 836 | 3 | n.a. | 77 | 1,096 | 168 | 784 | 2 | 1,865 | 25 | 78 | n.a. |
| | 2017 | 972 | 3 | n.a. | 32 | 1,488 | 333 | 733 | 6 | 2,000 | 25 | 95 | n.a. |
| | 2018 | 763 | 3 | n.a. | 40 | 1,475 | 219 | 798 | 10 | 2,121 | 25 | 100 | n.a. |
| | 2019 | 716 | 319 | n.a. | 41 | 1,402 | 232 | 1,175 | 33 | 3,127 | 27 | 96 | n.a. |
| | 2020 | 721 | 23 | n.a. | 41 | 1,419 | 211 | 1,007 | 23 | 4,129 | 28 | 327 | n.a. |
| 2021 | 714 | 3 | n.a. | 46 | 1,338 | 232 | 1,003 | 32 | 3,135 | 91 | 371 | n.a. | |
| 2022 | 719 | 2 | n.a. | 35 | 1,268 | 227 | 1,267 | 41 | 2,858 | 198 | 102 | n.a. | |

[注 1] 国際収支関連統計の基準変更*により、2013年以前と2014年以降のデータに連続性はない。→表中2013年以前の数値をグレー色・イタリックで表示。

[注 2] 「n.a.」は not applicable 公開情報なし、「△」は引き上げ超過を示す。「0」は単位未満、「-」は実績なしを示す。

出所：JETRO 直接投資統計 (<https://www.jetro.go.jp/world/japan/stats/fdi.html>) より作成。

4.2.3. 特許関連のデータ

調査対象国の「日本→対象国」の特許出願件数の推移を、表 23 に示す。なお、調査対象国の特許出願件数の推移の把握にあたって、「日本→対象国」の出願件数は、2021 年データでポルトガル、イタリア、2018～2020 年及び 2013 年以前のデータでインドネシアのデータの欠損が見られた。

国別の特徴では、特許出願件数が毎年一定程度見られる国（出願件数の増減が少ない国）が、メキシコ、インドネシア、マレーシア、ベトナム、タイ、サウジアラビア、インド等が挙げられる。

直近 10 年の平均の特許出願件数（括弧内は平均件数）では、スペイン（13 件）、メキシコ（834 件）、ポルトガル（1 件）、ブラジル（1322 件）、イタリア（99 件）、インドネシア（2098 件）、マレーシア（1242 件）、ベトナム（1053 件）、タイ（2276 件）、サウジアラビア（53 件）、インド（4154 件）、トルコ（29 件）であった。

表 23 特許出願件数（「日本→対象国」）

| | スペイン | メキシコ | ポルトガル | ブラジル | イタリア | インドネシア | マレーシア | ベトナム | タイ | サウジアラビア | インド | トルコ |
|------|------|-------|-------|-------|------|--------|-------|-------|-------|---------|-------|-----|
| 2007 | 9 | 498 | — | | 113 | | | | 25 | — | | 12 |
| 2008 | 37 | 630 | 4 | | | | 744 | | | | | 14 |
| 2009 | 31 | 632 | | | 93 | | | | | | | 7 |
| 2010 | 27 | 742 | | 1,826 | | | | 823 | 430 | 38 | | |
| 2011 | 21 | 759 | 9 | | 97 | — | 944 | 887 | 1,437 | 47 | 5,048 | 13 |
| 2012 | 19 | 992 | — | — | 124 | n.a. | 1,248 | 1,212 | 2,595 | n.a. | 6,221 | 20 |
| 2013 | 31 | 666 | 1 | 204 | 152 | n.a. | 528 | 368 | 649 | 7 | 326 | 9 |
| 2014 | 11 | 943 | 1 | 2,229 | 166 | 2,382 | 1,481 | n.a. | 3,061 | 9 | 5,338 | — |
| 2015 | 17 | 1,031 | 1 | 2,143 | n.a. | 2,548 | 1,420 | n.a. | n.a. | 108 | 4,857 | 77 |
| 2016 | 10 | 566 | n.a. | 311 | 163 | 1,053 | 748 | 466 | 1,141 | 45 | 962 | 27 |
| 2017 | 12 | 1,274 | n.a. | 1,717 | n.a. | 2,407 | 1,553 | 1,373 | 3,371 | 104 | 4,490 | 34 |
| 2018 | 13 | 615 | n.a. | 931 | 2 | n.a. | 1,127 | 769 | 2,556 | 33 | 1,946 | 38 |
| 2019 | 9 | 751 | n.a. | 1,125 | 26 | n.a. | 961 | 967 | 2,040 | 23 | 3,357 | 10 |
| 2020 | 4 | 659 | n.a. | 1,985 | 60 | n.a. | 1,930 | 1,645 | 2,181 | 33 | 4,381 | 21 |
| 2021 | 3 | 844 | n.a. | 1,257 | n.a. | 2,101 | 1,427 | 1,620 | 2,921 | 115 | 4,617 | 28 |

〔注 1〕 表中の「—」は、当該国からの報告、又は該当する件数がないことを示す（特許庁）

〔注 2〕 (n.a.) not applicable。特許庁、WIPO 等での公開情報なし。

〔注 3〕 空欄は、WIPO データにてデータ欠損値。

出所：各年の特許庁「特許行政年次報告書」（WIPO IP Statistics Data Center を基に特許庁作成）

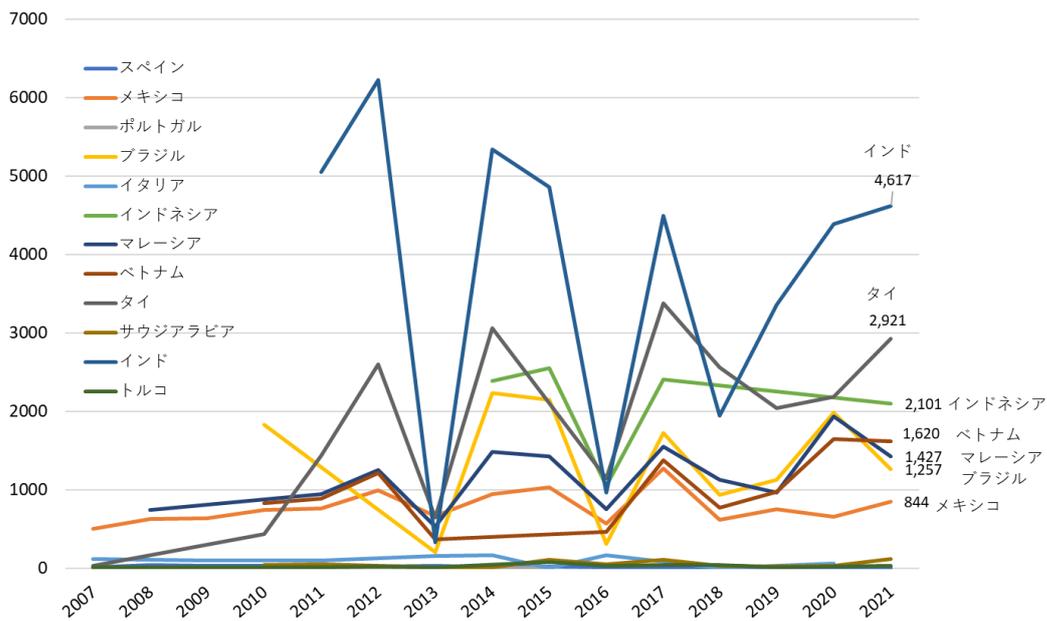


図 21 特許出願件数（日本→対象国）の推移（2007～2022年）（単位：件数）

出所：各年の特許庁「特許行政年次報告書」（WIPO IP Statistics Data Center を基に特許庁作成足）⁶⁰

調査対象国の「対象国→日本」の特許出願件数の推移は、表 24 の通りである。併せて、図 22 にて経年の推移を示す。なお、「対象国→日本」の出願件数の推移については、データ欠損の少ない、特許庁行政年次報告書（日本における国籍・地域別出願件数表）の値を用いて補足した。

国別の特徴では、特許出願件数が毎年一定程度見られる国（出願件数の増減が少ない国）として、スペイン、イタリア、インド等が挙げられる。

直近 10 年の平均の特許出願件数（括弧内は平均件数）では、スペイン（262 件）、メキシコ（30 件）、ポルトガル（27 件）、ブラジル（65 件）、イタリア（814 件）、インドネシア（2 件）、マレーシア（33 件）、ベトナム（7 件）、タイ（81 件）、サウジアラビア（98 件）、インド（236 件）、トルコ（56 件）であった。

⁶⁰ データについては、特許行政年次報告書（2009 年版～2023 年版）、主要国・地域・機関における 20xx 年の居住国・地域別出願件数表（特許）（2015 年版～2023 年版）より作成（<https://www.jpo.go.jp/resources/report/nenji/index.html>）。また、2014 年版以前については、以下を参照した

（<https://warp.da.ndl.go.jp/info:ndljp/pid/11239397/www.jpo.go.jp/shiryou/toukei/gyosenenji/index.html>）

表 24 特許出願件数（「対象国→日本」）

| | スペイン | メキシコ | ポルトガル | ブラジル | イタリア | インドネシア | マレーシア | ベトナム | タイ | サウジアラビア | インド | トルコ |
|------|------|------|-------|------|------|--------|-------|------|------|---------|-----|------|
| 2007 | 242 | | | | 818 | | | | | | 140 | |
| 2008 | 257 | | | | 820 | | | | | | 214 | |
| 2009 | 218 | | | | 683 | | | | | | 135 | |
| 2010 | 240 | | | 82 | 733 | | | | | | 162 | |
| 2011 | 226 | 34 | | 67 | 753 | | 27 | | 9 | | 170 | 26 |
| 2012 | 264 | 30 | n.a. | 74 | 688 | n.a. | 28 | n.a. | 17 | n.a. | 235 | 36 |
| 2013 | 146 | n.a. | n.a. | 49 | 751 | n.a. | n.a. | n.a. | n.a. | n.a. | 122 | 20 |
| 2014 | 242 | 19 | 20 | 88 | 757 | n.a. | 37 | n.a. | 51 | 122 | 228 | n.a. |
| 2015 | 243 | 33 | 23 | 75 | 765 | 8 | 38 | n.a. | n.a. | 79 | 235 | 30 |
| 2016 | 189 | n.a. | n.a. | 36 | 572 | 3 | n.a. | n.a. | 15 | n.a. | 149 | 37 |
| 2017 | 259 | 28 | 29 | 68 | 873 | — | 31 | n.a. | n.a. | 60 | 225 | 67 |
| 2018 | 166 | n.a. | n.a. | 29 | 527 | 3 | n.a. | n.a. | 31 | n.a. | 123 | 27 |
| 2019 | 125 | n.a. | n.a. | 25 | 479 | — | n.a. | n.a. | 46 | n.a. | 141 | 22 |
| 2020 | 142 | 28 | 32 | 28 | 549 | — | n.a. | n.a. | 41 | n.a. | 146 | 27 |
| 2021 | 295 | 23 | 34 | 48 | 839 | 1 | 31 | 6 | 118 | 80 | 233 | 93 |
| 2022 | 324 | 27 | 39 | 68 | 870 | 2 | 26 | 2 | 90 | 28 | 259 | 63 |

〔注 1〕 表中の「—」は、当該国からの報告、又は該当する件数がないことを示す（特許庁）

〔注 2〕 (n.a.) not applicable。特許庁、WIPO 等での公開情報なし。

〔注 3〕 空欄は、WIPO データにてデータ欠損値。

出所：各年の特許庁「特許行政年次報告書」（WIPO IP Statistics Data Center を基に特許庁作成；青字部分は特許庁データに基づく補足）

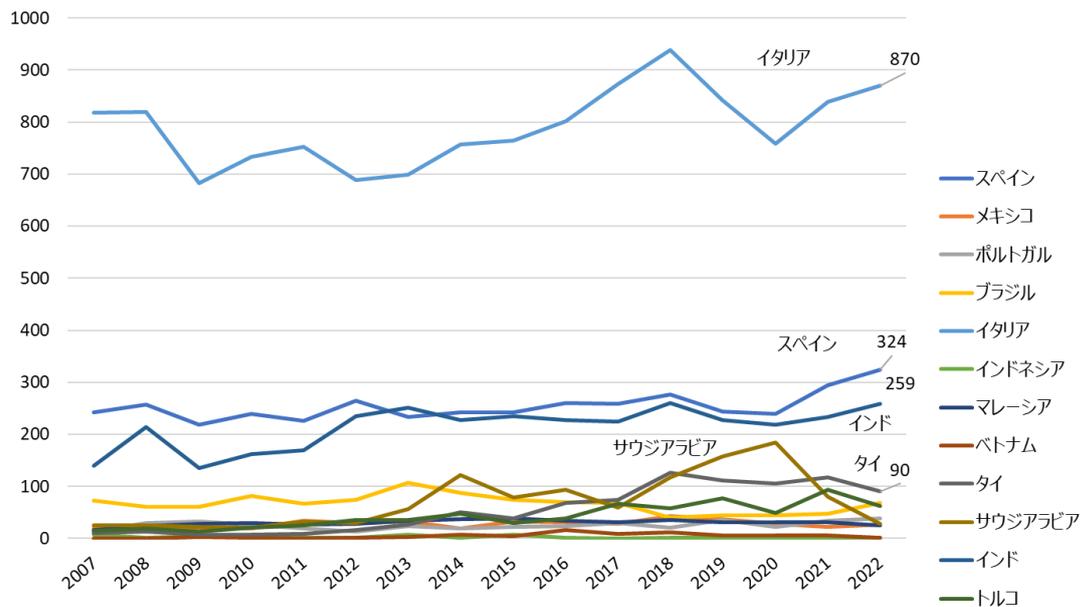


図 22 特許出願件数（対象国→日本）の推移（2007～2022年）（単位：件数）

出所：各年の特許庁「特許行政年次報告書」（WIPO IP Statistics Data Center を基に特許庁作成）⁶¹

⁶¹ データについては、特許行政年次報告書（2009年版～2023年版）、主要国・地域・機関における20xx年の居住国・地域別出願件数表（特許）（2015年版～2023年版）より作成（<https://www.jpo.go.jp/resources/report/nenji/index.html>）。また、2014年版以前については、以下を参照した

（<https://warp.da.ndl.go.jp/info:ndljp/pid/11239397/www.jpo.go.jp/shiryou/toukei/gyosenenji/index.html>）

PPH申請件数については、表25、表26に示す。収録データは2011年以降である。PPH申請件数のうち、「第1庁：日本・第2庁：対象国」の場合、申請数が多い国はメキシコ、インドネシア、マレーシア、タイ等の一部の国に限られている。

「第1庁：日本・第2庁：対象国」の直近10年の平均のPPH申請件数で経年的に一定の申請件数がある国（括弧内は平均件数）は、メキシコ（101件）、ブラジル（67件）、インドネシア（194件）、マレーシア（161件）、タイ（363件）であった。

表25 PPH申請件数（第1庁：日本、第2庁：対象国）（単位：件数）

| | スペイン | メキシコ | ポルトガル | ブラジル | イタリア | インドネシア | マレーシア | ベトナム | タイ | サウジアラビア | インド | トルコ |
|------|------|------|-------|------|------|--------|-------|------|-----|---------|-----|-----|
| 2011 | | 1 | | | | | | | | | | |
| 2012 | 1 | 21 | | | | | | | | | | |
| 2013 | | 39 | | | | 25 | | | | | | |
| 2014 | | 50 | | | | | 17 | | 108 | | | |
| 2015 | 0 | 85 | | | | 265 | 122 | | 168 | | | |
| 2016 | 1 | 111 | 0 | | | 137 | 166 | 0 | 390 | | | |
| 2017 | 0 | 90 | 0 | | | 0 | 178 | 0 | 455 | | | |
| 2018 | 5 | 200 | 0 | 41 | | 251 | 247 | 0 | 463 | | | |
| 2019 | 1 | 169 | 0 | 83 | | 373 | 252 | 0 | 545 | | | 0 |
| 2020 | 0 | 101 | 0 | 71 | | 112 | 224 | 0 | 411 | 0 | 0 | 2 |
| 2021 | 0 | 78 | 0 | 66 | | 0 | 84 | 91 | 348 | 0 | 0 | 0 |
| 2022 | 0 | 89 | 0 | 72 | | | 0 | | 381 | | | 1 |

〔注〕特許庁 PPH ポータル「統計情報」では、「統計情報の提供がない知財庁、集計期間中に PPH 申請が 0 件であった知財庁の統計情報は掲載されていません。」と記述されている。「PPH 申請件数データにて、①「国名あり、PPH 申請件数なし」は原則、PPH 申請件数 0 と見做して上表に記入。②「国名なし」は空欄とした。

出所：特許庁 PPH ポータル PPH 申請件数（2011 年～2022 年）より作成。

〈<https://www.jpo.go.jp/toppage/pph-portal-j/data/requests.xlsx>〉

表26 PPH申請件数（第1庁：対象国、第2庁：日本）（単位：件数）

| | スペイン | メキシコ | ポルトガル | ブラジル | イタリア | インドネシア | マレーシア | ベトナム | タイ | サウジアラビア | インド | トルコ |
|------|------|------|-------|------|------|--------|-------|------|----|---------|-----|-----|
| 2011 | | | | | | | | | | | | |
| 2012 | 2 | | | | | | | | | | | |
| 2013 | 0 | | | | | | | | | | | |
| 2014 | 0 | | | | | 1 | | | | | | |
| 2015 | 5 | | | | | 0 | | | | | | |
| 2016 | 0 | 1 | | | | 0 | | | | | | |
| 2017 | 2 | 0 | | | | 1 | | | | | | |
| 2018 | 1 | 1 | | | | 0 | | | | | | |
| 2019 | 0 | 1 | | | | 0 | | | | | | |
| 2020 | 0 | 0 | 0 | 1 | | 0 | 0 | 0 | 0 | 0 | 6 | 0 |
| 2021 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 5 | 0 |
| 2022 | 5 | 3 | 0 | 1 | | 1 | | | | | 4 | |

〔注〕特許庁 PPH ポータル「統計情報」では、「統計情報の提供がない知財庁、集計期間中に PPH 申請が 0 件であった知財庁の統計情報は掲載されていません。」と記述されている。「PPH 申請件数データにて、①「国名あり、PPH 申請件数なし」は原則、PPH 申請件数 0 と見做して上表に記入。②「国名なし」は空欄とした。

出所：特許庁 PPH ポータル PPH 申請件数（2011 年～2022 年）より作成。

〈<https://www.jpo.go.jp/toppage/pph-portal-j/data/requests.xlsx>〉

>

日本から対象国という方向での特許出願件数・PPH 申請件数は、日本企業等のビジネス活動における対象国への関心の程度、対象国から日本という方向での特許出願件数・PPH 申請件数は、対象国企業等のビジネス活動における日本への関心の程度の指標となる。前者の件数が高い国は、国内ユーザーに向けた現地語から日本語への機械翻訳サービスの需要が高いと見込まれ、後者の件数が高い国は、現地ユーザーに向けた日本語から現地語への機械翻訳サービスの需要が高いと見込まれる。

4.3. その他の経済的データについて

その他の経済的データは、調査対象国別に、(1) 当該国の概況、(2) 日本企業の進出状況、ビジネス展開状況、(3) 日本との関係、(4) 当該国の製造業の状況、(5) 地理的・地政学的重要性を示す。

4.3.1. スペイン

(1) 当該国の概況

現政権は、2023 年 11 月に発足した、サンチェス社会労働者党 (PSOE) による政権であり、首相は、ペドロ・サンチェス氏である。経済に係る課題は、財政安定化と福祉国家の再構築、経済成長の強化、質の高い雇用の創出を掲げていたが、COVID-19 の影響を受け、欧州復興基金を活用し、2021～2023 年にかけて 720 億ドルを投入し、「復興・改革・強靱化計画」を進めている。COVID-19 の拡大、ロシアのウクライナ侵攻等を経て、観光セクターの回復や電力価格の低下等も踏まえ、2023 年の経済成長率の予測は 2.3% に上方修正されたものの、失業率は 12.9% と EU 内でも高い値となっている⁶²。

(2) 日本企業の進出状況、ビジネス展開状況

日本企業の進出数は、2022 年 10 月現在、423 社であり、2022 年から既存進出企業の再投資が活発化している。AGC は設備増強 (投資総額、約 120 億円) を決定、花王は合成香料生産設備を増設した。また、NTT データが南部マラガに拠点を開設している。M&A ではリニューアブル・ジャパンが太陽光発電所 2 カ所を取得し、オリックスは再生可能エネルギー開発のエラワンエナジーを完全子会社化した。その他に、東芝三菱電機産業システムは港湾・ターミナル向け自動化事業を買収している⁶³。

⁶² 外務省「国・地域」情報：スペイン王国
(<https://www.mofa.go.jp/mofaj/area/spain/data.html>)

⁶³ JETRO (2023)「スペインの貿易と投資」, 2023 年 10 月.

表 27 ス페인への日本企業の進出状況

| 項目 | 主な内容 |
|----------|---|
| 日系企業の進出数 | 423 社 (2022 年 10 月現在) ⁶⁴ |
| 進出状況 | <ul style="list-style-type: none"> 進出企業の再投資が活発化 (AGC、花王等の設備増強、NTT データの拠点開設、M&A でリニューアブル・ジャパンは太陽光発電所の取得、オリックスは、エラワンエナジーの完全子会社化を図った) |
| 在留邦人数 | 9,712 人 (2023 年 10 月現在) ⁶⁵ |

(3) 日本との関係

2018 年の外交関係樹立 150 周年を機に、日本・スペイン・シンポジウム等を通じ様々な分野で関係が活性化している。2018 年 10 月には、首脳間で「戦略的パートナーシップに関する日本国政府とスペイン王国政府との間の共同声明」が公表された。経済分野における協力は、官民の経済産業関係を強化する決意を表明し、これまでに実施してきた、日本貿易振興機構 (JETRO) とスペイン貿易投資庁 (ICEX) の間の、第三国における日本企業とスペイン企業の協力促進のための情報共有、それぞれの相手国における投資を促進するためのビジネス界に対する相互補助及びスタートアップの発展の奨励を含む、両国貿易・投資の円滑化のための協力を歓迎した。2022 年には、岸田首相は、NATO 首脳会合 (マドリード開催) に初めて出席し、再生可能エネルギー、デジタルなどの分野での企業間協力やウクライナやインド太平洋を含む地域情勢の意見を交換した^{66,67}。

(4) 当該国の製造業の状況

スペインの 2022 年の対内直接投資 (届け出ベース、フロー) は前年比 6.4%増の 250 億 8,800 万ユーロであり、業種別で最大の運輸・通信分野 (63 億 5,600 万ユーロ) は、5G 整備に向けた国外投資会社との提携が加速している。プログラミング・IT サービスは、IT バブル期の 2000 年以来最高の 17 億 4,300 万ユーロに達した。米 IT 大手によるデータセンター開設の動きも活発で、アマゾン ウェブ サービス(AWS)やグーグルが稼働開始し、メタの設置計画も報じられている。製造業と不動産・企業向けサービスは、前年からそれぞれ 4 倍、2.8 倍の対内直接投資額が大幅増加した。製造業の大型案件は、独シーメンス・エナジーによる風力発電機器製造シーメンス・ガメサ・リニューアブルエナジーの株式公開買い付けで、32 億ユーロを投げ持ち株比率を 92.7%に引き上げた。電気・ガス・水道・環境については、20 億 4,800 万ユーロで前年比 65.8%減だった。再生可能エネルギー投資は引き揚げもあり、59.3%減の 14 億 5,200 万ユーロであった。エネルギー大手による独立発電事業者の買収は引き続き活発で、仏電力大手エンギーとクレディ・アグリコル・アシュアラ

⁶⁴ 外務省 (2022) 「海外進出日系企業拠点数調査」。

⁶⁵ 外務省 (2023) 「海外在留邦人数調査統計」。

⁶⁶ 前掲：外務省「国・地域」情報：スペイン王国

⁶⁷ 外務省 (2023) 「外交青書」, p.127.

ンスが再エネ大手エオリア・レノバブレスの株式 97.3%を取得した。一方で、国内エネルギー大手によるエネルギー移行資金調達のための資産売却も活発で、レプソルは世界全体の再エネ事業と上流事業の持ち分 25%をフランスとスイスのエネルギー投資会社などに合計 57 億ユーロで譲渡した⁶⁸。

(5) 地理的・地政学的重要性

① 当該国と日本との友好関係

スペインは、欧州とアフリカを結ぶ地理的な要所に位置し、地中海地域との結びつきも強い。日本とは特に EU との連携を通じて、共通の国際的な課題に対処し協力している。経済的相互補完性については、産業分野は技術力や資源の相違等により相互に補完的で、製造業やサービス分野での貿易が多い。日本からは自動車や鉄鋼、自動二輪車、日本へは豚肉、自動車、石油精製品などが主要品目となっている。企業間の相互投資が進み、ビジネス環境が向上している。観光業も拡大し文化的な交流が経済に寄与している。再生可能エネルギー分野（洋上風力等）でも協力が進んでいる。

② 経済安全保障やその他リスクに関する状況

日本とスペインの両国間で経済安全保障に係る具体的な事項はない。スペインにおけるリスクについて、世界経済フォーラムの「エグゼクティブ・オピニオン調査 (EOS)」は、急激な物価変動・長期化するインフレ、主要経済国の累積債務危機、生活費危機、費用や生活破綻（生活苦）の危機、戦略資源の地政学的な覇権争いと不法な取引・経済活動を挙げた⁶⁹。

4.3.2. メキシコ

(1) 当該国の概況

現政権は、2018 年 12 月 1 日にアンドレス・マヌエル・ロペス・オブラドール大統領が就任した（任期 6 年）。ロペス・オブラドール大統領は、既成政治を否定し、一般大衆を優先する政治を行うため、大統領の不逮捕特権の廃止、歴代大統領の年金廃止、高級官僚の給与引き下げ、最低賃金引き上げ、高齢者・障害者の年金額引き上げ、若年層の雇用促進、南部を中心としたインフラ計画、先住民・貧困層への優遇政策、反増税・反債務増加、新規製油所の建設、新たに創設された国家警備隊による治安維持・改善等の政策を掲げている⁷⁰。2022 年のメキシコの実質 GDP 成長率は 3.0%と堅調な成長を維持するとともに、内需は回復し民間消費支出は通年で 6.1%の成長を記録した。他方、米国向けの輸出品製造の活

⁶⁸ 前掲：JETRO（2023）「スペインの貿易と投資」, 2023 年 10 月。

⁶⁹ World Economic Forum（2023）「第 18 回グローバルリスク報告書 2023 年版」。

⁷⁰ 外務省「国・地域」情報：メキシコ合衆国
〈<https://www.mofa.go.jp/mofaj/area/mexico/data.html>〉

発化により、中間財・資本財輸入が増え、貿易収支は赤字基調となっている⁷¹。

(2) 日本企業の進出状況、ビジネス展開状況

日本企業の進出数は、2022年10月現在、1312社であり、2021/2022年は、豊田合成、JFE商事、テイ・エステック、川崎重工、東レ、エフテック、シークス、サカタのタネ、横浜ゴム、三井物産、ダイキン工業、住江織物、ユーラスエナジーホールディングス、フジテック等が進出している。対日関係については、対日輸入（日本→メキシコ）は、構成比で31.5%を占める電気・電子機器は前年比6.4%減の57億6,900万ドルであり、特にリチウム電池を含む蓄電池が82.8%減となった。一方で、中国からの蓄電池の輸入は59.2%増で、市場競争が激化している⁷²。

表 28 メキシコへの日本企業の進出状況

| 項目 | 主な内容 |
|----------|---|
| 日系企業の進出数 | 1,312社（2022年10月現在） ⁷³ |
| 進出状況 | <ul style="list-style-type: none"> 近年の対内直接投資案件として、豊田合成はエアバックを生産し、JFE商事は鋼板の加工・販売を、川崎重工は4輪車の製造を、ダイキン工業は小型エアコン、インバーターの製造を行うため新工場を建設している。また、サカタのタネは、メキシコ及び中南米市場向けの野菜の品種開発・マーケティングを行っている。それ以外では、ユーラスエナジーホールディングスは風力や太陽光発電事業を実施し、フジテックはエレベータ企業を買収した。 |
| 在留邦人数 | 10,253人（2023年10月現在） ⁷⁴ |

(3) 日本との関係

外交関係は、1888年に樹立され、2018年には日墨外交関係樹立130周年を迎え、2022年は日本人メキシコ移住（1897年）から125周年を記念し、祝賀行事が開催された。近年は、経済関係が強化され、中南米地域で最多の約1,300社の日系企業が進出している。

2022年に、日本－メキシコ外相会談（林－エブラル外相）を実施し、メキシコとの戦略的連携の一層の強化を確認したほか、国連全体の機能強化を含め、国際場裡における緊密な連携を確認した。経済面では、ビジネス環境の整備、法的安定性の重要性、環太平洋パートナーシップに関する包括的及び先進的協定（CPTPP）のハイスタンダードの維持を確認し、メキシコが議長国を務める太平洋同盟と日本の間で協力を進めることで一致した⁷⁵。

⁷¹ JETRO（2023）「メキシコの貿易と投資」
 〈https://www.jetro.go.jp/world/cs_america/mx/gtir.html〉

⁷² 前掲：JETRO（2023）「メキシコの貿易と投資」

⁷³ 前掲：外務省（2022）「海外進出日系企業拠点数調査」

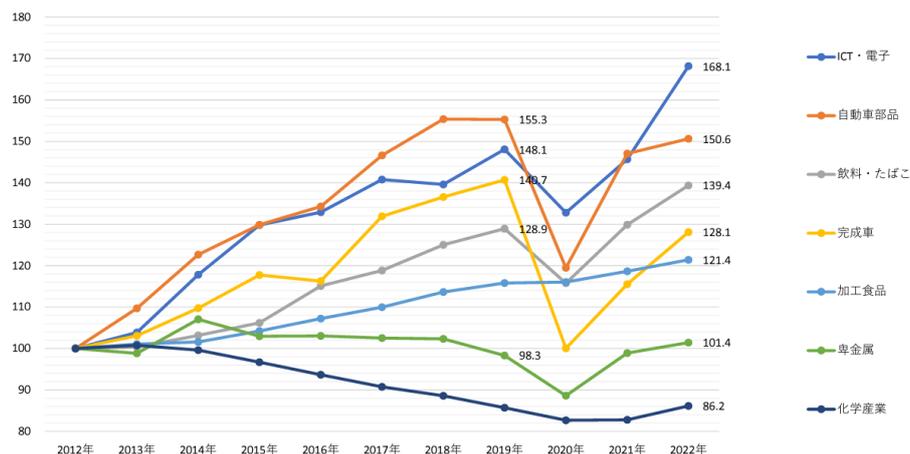
⁷⁴ 前掲：外務省（2023）「海外在留邦人数調査統計」

⁷⁵ 前掲：外務省（2023）「外交青書」, pp.114-116.

(4) 当該国の製造業の状況

メキシコの 2022 年の対内直接投資（国際収支ベース、ネット、フロー）は、前年比 14.8%増の 362 億 1,500 万ユーロドルであり、2013 年に次ぐ過去 2 番目の投資額を記録した。2022 年の対内投資額の内訳をみると、新規投資が全体の 49.6%を占める。業種別で最も構成比の大きい製造業への投資は 13,710 百万ドルで全体の約 38%を占める。特に好調な部門は、自動車産業等である^{76,77}。

製造業付加価値の 5%以上を占める主要分野について、粗付加価値（実質値）を 2012 年 =100 として指数化し、過去 10 年間の推移をみると ICT・電子機器製造の 2022 年の水準は 168.1 に達し、新型コロナ前を 20 ポイントも上回った。自動車部品製造は 150.6 と危機前の 155.3 とほぼ同じ水準、完成車製造は 128.1 で、危機前の水準に戻るには一定の時間が必要とされる。生産が好調な分野の背景には、対米輸出を視野に入れたニアショアリング（近隣国へのアウトソーシング）の観点から、中国などからメキシコへの生産移転の影響がみられる⁷⁸。



(注) 製造業付加価値の5%以上を占める業種の各年1-9月の実質粗付加価値のデータを2012年1-9月を100として指数化。

(出所) 国立統計地理情報院 (INEGI)

図 23 主要製造業の粗付加価値指数の推移（2012年1-9月=100とした場合）

出所：JETRO ビジネス短信「第3四半期の産業別 GDP、製造業が前年同期比 7.3%増と好調」（2022年12月1日）」添付資料

⁷⁶ 前掲：JETRO（2023）「メキシコの貿易と投資」

⁷⁷ JETRO（2023）「第2四半期 GDP は好調持続」, JETRO ビジネス短信, 2023年9月5日。
 〈<https://www.jetro.go.jp/biznews/2023/09/62e6065ca3006ecc.html>〉

⁷⁸ JETRO（2022）「第3四半期の産業別 GDP、製造業が前年同期比 7.3%増と好調」, JETRO ビジネス短信, 2022年12月1日。

〈<https://www.jetro.go.jp/biznews/2022/12/73a4f7f6c9d75ec4.html>〉

*添付資料 〈https://www.jetro.go.jp/view_interface.php?blockId=34829652〉

(5) 地理的・地政学的重要性

①当該国と日本との友好関係

メキシコは、米国からみて中南米の入り口に当たり、約 3,000km にわたって国境線を接している地政学上の重要な位置を占める。1994 年の NAFTA 発効以降、米国との経済関係が強まり、輸入全体の約 44%、輸出全体の約 82%を米国が占める（2022 年墨中銀）。経済発展を進めつつある中進国とはいえ、中南米地域ではブラジルと並ぶ大国で、国際場裏において果たしている役割も小さなものではなくなりつつある⁷⁹。メキシコとの経済的相互補完性については、2005 年 4 月に「日墨経済連携強化のための協定」（EPA）が発効し、2018 年 12 月には日本とメキシコもメンバーである CPTPP が発効した。日本からは自動車部品、自動車、鉄鋼のフラットロール製品、音響・映像機器の部分品、金属加工機器等が輸出され、メキシコからは電気機器、科学光学機器、豚肉、一般機器、塩、果実、自動車、非鉄金属鉱等が輸入される。

②経済安全保障やその他リスクに関する状況

日本とメキシコの両国間で経済安全保障に係る具体的な事項はない。メキシコにおけるリスクについて、世界経済フォーラムの「エグゼクティブ・オピニオン調査（EOS）」によると、急激な物価変動・長期化するインフレ、不法な取引・経済活動、長期化する経済停滞、国家の崩壊、社会的幸福と結束の浸食等が懸念される⁸⁰。

4.3.3. ポルトガル

(1) 当該国の概況

現政権は、アントニオ・コスタ首相による社会党政権であり、2022 年予算法審議にて予算法案の否決を受け、共和国議会議員選挙及び欧州選挙区の再投票により、2022 年 3 月に第 3 次コスタ政権が発足し、第 2 次政権時に掲げた「気候変動」、「人口減少」、「デジタル移行」、「社会的不均衡の改善」を最重要課題として掲げた。一方で、同政権では、リチウム開発、水素製造、データセンター等のプロジェクトを巡り汚職疑惑がもたれ、2023 年 11 月 7 日にコスタ首相は辞意を表明し⁸¹、2024 年 3 月 10 日に総選挙が行われた。ポルトガルの経済は、輸出入の約 7 割、対ポルトガル直接投資の約 8 割を EU 域内国が占めている。また、市場拡大のため、ブラジル、アンゴラ等のポルトガル語圏諸国共同体（CPLP）⁸²、中

⁷⁹ 在メキシコ日本国大使館、「メキシコへの経済協力の方針」．〈<https://www.mx.emb-japan.go.jp/keizai/keizai-kyouryoku01.htm>〉

⁸⁰ 前掲：World Economic Forum（2023）「第 18 回グローバルリスク報告書 2023 年版」．

⁸¹ JETRO（2023）「ポルトガルの首相が辞任、2024 年 3 月に総選挙へ」, JETRO ビジネス短信, 2023 年 11 月 26 日.

⁸² ポルトガル語圏諸国共同体の加盟国は、全 9 か国（ポルトガル、ブラジル、アンゴラ、カーボベルデ、ギニアビサウ、モザンビーク、サントメ・プリンシペ、赤道ギニア、東テ

国を始めとするアジア、北米・中南米等との経済交流拡大を図っている⁸³。

ポルトガル政府では、EU 復興基金に基づき、「復興・再建計画 (Plano de Recuperação e Resiliência : PRR)」を 2021 年 4 月に発表し、EU 理事会で 7 月に承認された。同計画は、新型コロナウイルス感染拡大後の持続可能な経済成長を目的に、復興、気候中立の達成、デジタル化への移行を 3 本柱に 2026 年までの投資計画を推進している⁸⁴。

(2) 日本企業の進出状況、ビジネス展開状況

日本企業の進出数は、2022 年 10 月現在、114 社である。ポルトガルへの日系企業の進出は、1970～1980 年代を通じて、労働コストの優位性を背景に日本企業が進出したものの、欧州における工場等の拠点は、中東欧に移転が進んだ。ポルトガルに進出している企業は、自動車産業を中心に、トヨタ自動車、三菱ふそうトラック・バス、デンソー、オムロン、昭和電工、メッツ、内山工業、豊和繊維工業、矢崎総業等が生産拠点を置いている。また、根本特殊化学、信越化学工業、YKK、カゴメ、荒井商事等が事業を展開している。2018 年には、日立コンサルティング (画像診断のための研究開発センター)、コニカミノルタ (Office of Future 用イノベーションセンター)、帝人 (自動車部品メーカーの買収)、カゴメ (新製造ラインの建設)、トヨタ自動車 (CaetanoBus 社へ水素技術の提供) 等の企業が投資を行っている⁸⁵。

対日貿易主要品目の構成比は、対ポルトガル輸出では、輸送用機器 (41.8%)、鉄鋼 (15.4%)、一般機械 (10.6%)、電気機器 (7.1%)、ゴム製品 (3.9%) である。一方で、対ポルトガル輸入では、輸送用機器 (28.3%)、衣類及び同附属品 (18.8%)、果実及び野菜 (10.9%)、電気機器 (8.8%)、はき物 (4.8%) である。

ポルトガルでは、グリーン水素サプライチェーンの構築を目指し、2023 年 6 月に「グリーン水素プロジェクト」への助成が正式に決定した。同国政府はエネルギー移行や対内投資の促進のための行政手続きの迅速化、簡素化の一環として、3 月からグリーン水素製造プロジェクトの環境影響評価の義務を撤廃している。東芝エネルギーシステムが 2022 年 5 月、PEM 型 (固体高分子型) 電解装置の開発・製造・販売を行うフュージョンフュエル社 (アイルランド) と、欧州やオーストラリア市場での電解装置拡販に向けた協業を検討することで合意したほか、丸紅は 2023 年 3 月にフロエン社 (ガス配送会社) が新設した水素配送パイプラインを用いた、ポルトガル初のグリーン水素注入実証事業を開始した⁸⁶。

イモール) であり、旧植民地諸国との連携、経済関係の強化、人的交流の促進、ポルトガル語の国際語化を図っている。

⁸³ 前掲：外務省「国・地域」情報：ポルトガル共和国
(<https://www.mofa.go.jp/mofaj/area/portugal/data.html>)

⁸⁴ JETRO (2023)「ポルトガル概況」

⁸⁵ フランシスコ・ジャヴィエル・エステヴェス (2019)「予想を超えるポルトガルの魅力」, JOI, 2019 年 1 月。

(https://www.joi.or.jp/wp-content/uploads/2022/10/Mag_201901_10_mfaPRT.pdf)

⁸⁶ JETRO (2023)「急速に進むグリーン水素サプライチェーン構築(スペイン、ポルトガ

表 29 ポルトガルへの日本企業の進出状況

| 項目 | 主な内容 |
|----------|---|
| 日系企業の進出数 | 114社（2022年10月現在） ⁸⁷ |
| 進出状況 | <ul style="list-style-type: none"> 自動車産業を中心に、トヨタ自動車、三菱ふそうトラック・バス、デンソー、オムロン、昭和電工、メッツ、内山工業、豊和繊維工業、矢崎総業等が生産拠点を置いている 2023年6月に「グリーン水素プロジェクト」への助成が正式に決定し、水素エネルギー関連のプロジェクトに日系企業も参加している（東芝エネルギーシステム、丸紅） |
| 在留邦人数 | 872人（2023年10月現在） ⁸⁸ |

（3）日本との関係

2020年には日ポルトガル修好通商条約調印から160年、2023年には1543年の種子島への鉄砲伝来から480年にあたり、文化交流が活発化している。2023年9月の日・ポルトガル外相会談（上川外務大臣ーゴメス・クラヴィーニョ外務大臣）では、自由で開かれたインド太平洋の実現に向けて、ポルトガルとの協力の進展とともに、ポルトガルの欧州におけるGX、DXの中心となるための取組みと水素分野の具体的な協力の進展と発展を歓迎した⁸⁹。

（4）当該国の製造業の状況

ポルトガルの製造業の状況として、主要輸出品目である輸送用機器について、国別自動車生産台数は2022年に世界29位（322,404台、前年比11.2%プラス）となった。

（5）地理的・地政学的重要性

①当該国と日本との友好関係

ポルトガルは、欧州統合への積極的な参加、大西洋同盟を重視した立場をとっており、日本においても、自由で開かれたインド太平洋の実現に向けた協力関係の構築や、種子島への鉄砲伝来以降の480年に渡る交流の歴史から、文化交流が盛んに行われている。

②経済安全保障やその他リスクに関する状況

市場拡大のため、ブラジル、アンゴラ等のポルトガル語圏諸国共同体（CPLP）、中国を

ル)」、JETRO 地域・分析レポート、2023年7月20日。

〈<https://www.jetro.go.jp/biz/areareports/special/2023/0503/afff5453326bb72c.html>〉

⁸⁷ 前掲：外務省（2022）「海外進出日系企業拠点数調査」

⁸⁸ 前掲：外務省（2023）「海外在留邦人数調査統計」

⁸⁹ 外務省「日・ポルトガル外相会談」、2023年9月21日。

始めとするアジア、北米・中南米等との経済交流拡大を図っている。

2024年3月10日に実施された議会選挙（一院制）で、与党（社会党）が敗北し、中道右派「民主主義同盟（AD）」が与党を僅差で上回り、勝利を収めたため、2015年以外の政権交代であり、少数政権の樹立に向けた準備が進む⁹⁰。同選挙では、新興の極右政党「シェーガ」が政治不信を追い風に躍進し、「民主主義同盟」に対して協力を強くアピールしている。

ポルトガルにおけるリスクについて、世界経済フォーラムの「エグゼクティブ・オピニオン調査（EOS）」によると、急激な物価変動・長期化するインフレ、生活費危機、主要経済国の累積債務の危機、極端なコモディティショックや相場の乱高下、長期化する経済停滞が懸念される⁹¹。

4.3.4. ブラジル

(1) 当該国の概況

現ルーラ政権（第3次）は、2023年1月に発足した。2023年10月の大統領選挙にて、当時現職のボルソナロ大統領を破り、12年ぶりに大統領に就任した。2019年に発足したボルソナロ前政権は、経済の自由化を掲げ、社会・経済開発を重視する開発モデルからの脱却を目指し、省庁再編、経済関連政策の策定の一元化を図った。他方、COVID-19パンデミックにより、ブラジル経済の成長率は-3.28%に下落し、経済状況悪化し、基礎的財政収支の赤字、政府総債務残高が膨れ上がった。第3次ルーラ政権では、貧困層の所得の底上げ、雇用と生産の好循環に向けた開発モデルに向けた施策を展開している⁹²。経済関係では、“ブラジルクスト”（ブラジル国内に所在する企業が国内で生産活動を行うにあたって直面する追加的コスト）が実態的に発生している。背景には、税の種類が多さと、連邦、州、市のそれぞれに税が存在し、納税に費やす時間はOECD高所得国平均（年158.8時間）の約10倍の時間を要する。このため、新政権発足直後から、税の簡素化や公平で透明性の高い税制の実現に向けた取組みを実施している⁹³。

(2) 日本企業の進出状況、ビジネス展開状況

ブラジルは、メキシコと並ぶ中南米の日系企業の生産拠点である。日本企業の進出数は、

⁹⁰ ロイター（2024）「ポルトガル中道右派が政権樹立へ準備、極右は不安定化を警告」、2024年3月12日。〈<https://jp.reuters.com/world/KRS6CC4SDZIWXPDAECQIESHVCY-2024-03-12/>〉

⁹¹ 前掲：World Economic Forum（2023）「第18回グローバルリスク報告書2023年版」。

⁹² 河合沙織（2023）「ブラジルにおけるマクロ経済政策の動向と第3次ルーラ政権の課題」、ラテンアメリカ・レポート、Vol.40, No.1, pp.26-40, IDE-JETRO。

〈https://www.jstage.jst.go.jp/article/latinamericareport/40/1/40_26/_pdf/-char/ja〉

⁹³ 岩崎英明（2024）「ブラジル経済概観と財政・税制改革への取組について」、ファイナンス、2024年1月。〈https://www.mof.go.jp/public_relations/finance/202401/202401k.pdf〉

2022年10月現在、682社であり、自動車のほか、幅広い産業で市場ニーズが拡大し、2023年にJETROが実施した「海外進出日系企業実態調査」においても、他の国・地域と比べても事業が大きく拡大している⁹⁴。事業を拡大する業種は、自動車、再エネ、エネルギーマネジメント、農業、デジタル、医療、産業資機材、食品、その他消費財、情報通信、金融、サービス等と幅広い業種に跨る。先進主要国ではEV化の進展に伴い、バイオ燃料が利用可能なHV車の導入拡大も期待されており、幅広い部品需要の下支えも期待される。

日本からブラジルへの対内直接投資額（国際収支ベース、フロー）は、自動車産業などがけん引し、36.4%増加した。2022年3月にトヨタ自動車は、カローラ（セダン）を製造しているサンパウロ州内のインダイアツーバ工場向けに5,000万レアルの投資を発表したほか、同年4月には、日産自動車がリオ・デ・ジャネイロ州に位置するヘゼンジ工場の近代化などのため、2022年から2025年の間に13億レアルの投資を行う旨発表。2022年9月には、ブリヂストンがバイーア州カマサリ工場の生産能力増強のため、2億7,000万レアル（約5,432億ドル）の投資を発表。2023年11月には、日産自動車は、ラテンアメリカ市場の戦略的拡大を目指し、大規模投資計画（2025年までに約28億レアル（約5億6336万ドル）を発表するとともに、向こう10年にわたる戦略的投資を行い、累計で62億レアルを予定している。

表30 ブラジルへの日本企業の進出状況

| 項目 | 主な内容 |
|----------|---|
| 日系企業の進出数 | 682社（2022年10月現在） ⁹⁵ |
| 進出状況 | <ul style="list-style-type: none"> 進出企業の事業拡大の意向が高まっている（自動車、再エネ、エネルギーマネジメント、農業、デジタル、医療、産業資機材、食品、その他消費財、情報通信、金融、サービス等と幅広い業種に跨る） ブラジルへの対内直接投資は、自動車産業が牽引し、工場等への投資を活発化している |
| 在留邦人数 | 46,902人（2023年10月現在） ⁹⁶ |

（3）日本との関係

2008年に、ブラジル移住100周年「日本ブラジル交流年」を、2015年は日ブラジル外交関係樹立120周年を迎えた。ブラジルには、海外で最大の日系社会（約200万人）があり、活発な要人往来が行われてきた。2014年以降は、二国間関係を戦略的グローバル・パートナーシップと位置づけ、政治、経済、人的交流のみならず治安、防災、刑事司法、環境、教育等、幅広い分野で二国間協力が進展している。マルチの分野では安保理改革等で連携

⁹⁴ JETRO（2023）「海外進出日系企業実態調査 中南米編」, 2023年12月.

〈https://www.jetro.go.jp/ext_images/Reports/01/c9936f87d6a3530c/20230028.pdf〉

⁹⁵ 前掲：外務省（2022）「海外進出日系企業拠点数調査」

⁹⁶ 前掲：外務省（2023）「海外在留邦人数調査統計」

を進めている⁹⁷。

(4) 当該国の製造業の状況

ブラジルの 2022 年の対内直接投資額（国際収支ベース、フロー）は、前年比 10.7%増の 435 億 9,100 万ドルだった。業種別にみると、工業分野で最も構成比が大きい自動車・トレーラー・車体（構成比 13.0%）は 48.0%増加した。また、中国企業によるブラジル向け投資は、ブラジル中央銀行が公開している統計に表れてこない場合があり、中銀統計によると中国からの対内直接投資額は 3,500 万ドルにとどまる。一方で、企業ごとの投資計画の発表では、長城汽車は 2021 年 8 月にサンパウロ州イラセマポリス市にあるドイツ・ダイムラーの工場を買収しブラジルに進出し、電気およびハイブリッドのスポーツタイプ多目的車（SUV）、ピックアップトラックを中心に年間 10 万台の生産を目指し、中南米諸国向けの輸出拠点を目指している。投資額は 10 年間で 100 億レアル（約 20 億ドル）が見込まれている。電気自動車の BYD オートも、2022 年 10 月にバイーア州の 3 工場に約 30 億レアル（約 6 億ドル）の投資計画を発表している⁹⁸。

(5) 地理的・地政学的重要性

① 当該国と日本との友好関係

日本とブラジルは、2023 年に国連安全保障理事会の非常任理事国をともに務めていたことから、2023 年 9 月に両国の外相会談で、国連全体の機能強化に関連した連携の必要性を確認したほか、日伯戦略的経済パートナーシップ賢人会議、9 月に日本ブラジル経済合同委員会、領事当局間協議及び科学技術協力合同委員会を開催した。

2004 年の日伯首脳会議で、「日伯 21 世紀協議会」が設置され、2006 年 7 月に経済分野で実現可能な具体的な優先課題を選定し、日伯間の戦略的パートナーシップの再活性化に貢献することを目的に、「日伯戦略的経済パートナーシップ賢人会議」⁹⁹が設立された。これまでに 12 回の会合・特別会合が実施され、直近の 2023 年 4 月 10 日開催された「第 11 回会合」では、脱炭素社会の実現に向けて、ブラジルには、過去 50 年にわたり自動車燃料として活用されてきたバイオエタノールの原料となるサトウキビの生産拡大や用途多様化のポテンシャルがあることから、バイオエタノールに加え、第二世代バイオエタノール、バイオガス、持続可能な航空燃料（SAF）、水素といったバイオエタノールの誘導品も注目すべき資源、製品として期待された。また、「アマゾン地域に於ける生態系保全と社会発展に対する積極的な貢献」では、アマゾン基金やその他の資金枠組みへの参加・活用、森林保全に関する技術供与といった国レベルでの公的な支援に加え、民間セクターとして如何な

⁹⁷ 外務省「国・地域」情報：ブラジル〈<https://www.mofa.go.jp/mofaj/area/brazil/data.html>〉

⁹⁸ JETRO（2023）「ブラジルの貿易と投資」, 2023 年 9 月 4 日。
〈https://www.jetro.go.jp/world/cs_america/br/gtir.html〉

⁹⁹ 賢人会議のメンバーは、産業界メンバーで構成。

る連携・協力が可能を議論した¹⁰⁰。

②経済安全保障やその他リスクに関する状況

中南米では、2018年にメキシコ、2019年にアルゼンチン、2021年にペルー、チリ、ホンジュラス、2022年にコロンビア、ブラジルと左派政権が誕生し、2000年代の「ピンクの潮流」が再来したとも言われている。しかし、コロナ禍やウクライナ紛争を背景とした景気悪化や汚職などをめぐる政治混乱で、各国の政権運営が困難になっている。2023年は、中央銀行の利上げなどを背景とした経済減速による政治混乱や治安悪化、利上げによる米ドル建て債務拡大に伴う左派的財政出動への制約、アルゼンチン10月総選挙におけるフェルナンデス政権の苦戦ないしは失脚など、事業環境の不透明性が増すリスクが存在する。

ブラジルにおけるリスクについて、世界経済フォーラムの「エグゼクティブ・オピニオン調査（EOS）」は、急激な物価変動・長期化するインフレ、不法な取引や経済活動、地経学上の対立、極端なコモディティショックや相場の乱高下、雇用や生活破綻（生活苦）の危機を挙げた¹⁰¹。

4.3.5. イタリア

(1) 当該国の概況

2022年7月に、ウクライナ情勢の影響を踏まえた国内経済の支援策に対する連立与党メンバー（「五つ星運動(M5S)」）の一部政令案の反対、投票欠席を受けて、当時のドラギ首相は辞意表明をした。同年9月に上下両院議員の選挙の結果、「イタリアの同胞」を中心とする「中道右派連合」が過半数の議席を獲得し、10月に「イタリアの同胞」のメローニ党首によるメローニ政権が発足した。前ドラギ政権は、EU・NATO等に立脚した親欧州主義・大西洋主義の基本姿勢の元、環境問題や移民問題等、地中海特有の問題を共有する国々と関係強化を図っていた¹⁰²。経済面では、政府による建設奨励施策やEUの復興基金を活用した復興パッケージ「再興・回復のための国家計画（PNRR）」（2021年7月、EU承認）を利用した公共投資計画が実施され、企業景況感の回復の兆しが見える¹⁰³。

(2) 日本企業の進出状況、ビジネス展開状況

日本企業の進出数は、2022年10月現在、400社であり、日系企業によるイタリア企業買

¹⁰⁰ 日伯戦略的経済パートナーシップ賢人会議（2023）「第11回日伯戦略的経済パートナーシップ賢人会議に関する最終報告書」, 2023年4月10日開催。

〈<https://www.mofa.go.jp/mofaj/files/100494333.pdf>〉

¹⁰¹ World Economic Forum（2023）「第18回グローバルリスク報告書2023年版」。

¹⁰² 外務省「国・地域」情報：イタリア共和国

〈<https://www.mofa.go.jp/mofaj/area/italy/data.html>〉

¹⁰³ JETRO（2023）「イタリアの貿易と投資」, 2023年11月1日。

〈<https://www.jetro.go.jp/world/europe/it/gtir.html>〉

収等の活発に行われている（日立製作所によるフィンメカニカ社傘下の鉄道関連事業の買収（2015年～）、NTTデータによるバリュー社（ITベンダー）の買収（2011）、NTTドコモのボンジョルノ社（コンテンツ配信）の買収等、NTTグループによる買収がよく知られている）。その他、デンソー、三菱電機、住友重機械工業等もイタリア企業を買収、進出している）。2022年の対日貿易を見ると、日本からイタリアへの輸入を品目で最大のものは、輸送機器（24.1%）である。2番目は、機械（23.7%）であり、工作機械、建設機械の増加が寄与している。3番目は、金属製品（13.2%）であり、鉄・鋳鉄・鋼鉄・合金鉄が前年比で2.6倍と大幅に伸びた。一方で、イタリアから日本への出品で、最大のものは、繊維・衣料品・皮革製品（構成比 23.7%）で前年比 19.8%増である。2022年の日本からの直接投資受入額は4億2,600万ユーロである。日本電産は11月にイタリアの工作機械メーカー「パーマ社」および同関連会社9社の株式取得にかかる譲渡契約締結を発表した。ダイキン工業では、同年4月に、デュプロマティックMS社（油圧機器メーカー）を2億2,000万ユーロで買収し、子会社化を発表した。アドバンテストは同年6月、パワー半導体用試験装置のCREA社の買収について最終合意したと発表した¹⁰⁴。

表 31 イタリアへの日本企業の進出状況

| 項目 | 主な内容 |
|----------|---|
| 日系企業の進出数 | 400社（2022年10月現在） ¹⁰⁵ |
| 進出状況 | <ul style="list-style-type: none"> M&Aが活発化（鉄道事業、ITベンダー、コンテンツ配信事業者等を買収） |
| 在留邦人数 | 11,856人（2023年10月現在） ¹⁰⁶ |

（3）日本との関係

2016年には日伊国交150周年を迎え、我が国がG7議長国であったことから、各大臣間の交流が進んだ。2021年には、イタリアがG20の議長国を務め、2024年にはG7議長国を務めることから、同様に各大臣会合が開催された。

2023年1月には、日伊首脳会談において、両国関係を「戦略的パートナー」に格上げすることで一致し、水素、鉄度、素材、宇宙等の分野での協力を進展させることを歓迎した¹⁰⁷。

1989年以降、日伊ビジネス・グループ（Italy-Japan Business Group：IJBG）会合が毎年開催され、日伊経済関係の強化のための活動を行っている。これまでに33回の会合が開催された。

¹⁰⁴ 前掲：JETRO（2023）「イタリアの貿易と投資」。

¹⁰⁵ 前掲：外務省（2022）「海外進出日系企業拠点数調査」

¹⁰⁶ 前掲：外務省（2023）「海外在留邦人数調査統計」

¹⁰⁷ 外務省（2023）「日伊首脳会談及びワーキング・ランチ」, 2023年1月10日。

〈https://www.mofa.go.jp/mofaj/erp/we/it/page4_005746.html〉

(4) 当該国の製造業の状況

イタリアの 2022 年の対内直接投資額（国際収支ベース、ネット、フロー）は、前年比 81.8%増の 292 億 3,100 万ユーロとなった。主要な投資元である EU のうち、ベルギーが 3 億 3,200 万ユーロから 113 億 8,000 万ユーロへ大幅増となり、前年引き揚げ超過だったドイツ、アイルランド、スウェーデンが流入超過へ転じたことなどが影響した。その他、米国からの投資が拡大した。

主な対内投資案件をみると、米マイクロソフトの大手ソリューション・プロバイダーであるアバナードが 2 月、北部エミリア・ロマーニャ州ボローニャに新しい欧州製造技術イノベーションハブを開設した。製造企業のデジタルトランスフォーメーションを支援していく。スペインの洋上風力発電開発企業ブルーフロート・エナジーは 4 月、イタリアの再生可能エネルギー企業ファルク・リニューアブルズと提携し、イタリアの 3 州で 5 つの浮体式洋上風力発電所を設置すると発表した。2023 年に入ってから、フランスの電気自動車（EV）用のバッテリーメーカーのオートモーティブ・セルズ・カンパニーが 3 月、中部モリーゼ州テルモリにリチウムイオンバッテリーの大規模生産工場を建設する計画を発表した¹⁰⁸。

(5) 地理的・地政学的重要性

① 当該国と日本との友好関係

イタリアとは、「戦略的パートナー」に格上げ以降、2024 年 4 月には「社会保障に関する日本国とイタリア共和国との間の協定（日・イタリア社会保障協定）」が発効され、企業及び駐在員等の負担（年金制度等への社会保障費の二重払いの負担）が軽減され、日・イタリア両国間の人的・経済交流が一層促進される可能性がある。2023 年 12 月には、半導体、AI 等の先端分野で協力する共同声明に署名し、経済関係において官民一体となった取組みが進みつつある¹⁰⁹。

② 経済安全保障やその他リスクに関する状況

イタリアは、2019 年に G7 主要国として唯一、中国の「一帯一路構想」に参画したが、メローニ政権が発足し、2023 年 12 月 6 日に離脱を正式に通知した。

2022 年 12 月に日本・イタリア・英国 3 国による次期戦闘機共同開発協力（「グローバル戦闘航空プログラムに関する共同首脳声明」）を発表した。2023 年 12 月には、「グローバル戦闘航空プログラム（GCAP）政府間機関の設立に関する条約」が署名され、GCAP 政府間機関（GIGO）の地位、任務と、締約国の義務等を定める¹¹⁰。

イタリアにおけるリスクについて、世界経済フォーラムの「エグゼクティブ・オピニオ

¹⁰⁸ 前掲：JETRO（2023）「イタリアの貿易と投資」, 2023 年 11 月 1 日。

¹⁰⁹ 経済産業省「共同声明」, <<https://www.meti.go.jp/press/2023/12/20231212003/20231212003-2.pdf>>

¹¹⁰ 外務省（2023）「『グローバル戦闘航空プログラム（GCAP）政府間機関の設立に関する条約』の署名」, <https://www.mofa.go.jp/mofaj/press/release/pressit_000001_00090.html>

ン調査（EOS）」は、主要経済国の累積債務危機、国家間の紛争、急激な物価変動・長期化するインフレ、気候変動への対応策の失敗、主要経済国の資産バブルを挙げた¹¹¹。

4.3.6. インドネシア

(1) 当該国の概況

ジョコ・ウィドド大統領政権は、2019年に再任され、2期目である。ジョコ政権では、①インフラ開発、②人材開発、③投資促進、④官僚改革、⑤適切な国家予算の執行を優先課題に取り組み、経済・社会政策では、鉄道、港湾、電力・エネルギー等のインフラ整備及び社会保障の充実を目標に掲げている。インドネシアは、マラッカ海峡等のシーレーン上の要衝に位置し、東南アジア諸国連合（ASEAN）の中で唯一のG20メンバー国であり、地域・国際社会の諸課題にイニシアティブを発揮している¹¹²。経済状況は、2022年の実質GDP成長率は5.3%で前年から1.6ポイント上昇した。資源価格の高騰などによって鉱物性燃料などの輸出額が伸びたことに加え、新型コロナウイルス（以下、新型コロナ）の影響で低迷していた消費や投資も緩やかに回復した¹¹³。

(2) 日本企業の進出状況、ビジネス展開状況

日本企業の進出数は、2022年10月現在、2,103社であり、トヨタ自動車、ダイハツ工業、三菱自動車、ホンダ技研、スズキ、ヤマハ、パナソニック、シャープ、エプソン、東レ、マンダム、ユニチャーム、ライオン、花王、味の素、ヤクルト、日清食品、旭硝子、公文等が進出している。海外進出日系企業拠点調査では、当該国への投資・進出の特長として、市場の成長性、現在の市場規模、人材の雇いやすさ（一般ワーカー、スタッフレベル）が挙げられた。また、問題点として、税制・税務手続きの効率性、法制度の整備状況（外資優遇・規制など）、行政手続きの効率性（許認可など）、人件費の水準、制度・政策の運用の透明性（産業政策、エネルギー政策、外資規制等）等が挙げられる¹¹⁴。

2022年の日本からインドネシアへの輸出は、150億3,200万ドルでほとんどの分野で増加した。主要な輸出品は、一般機器・原子炉・ボイラー、輸送用機器（鉄道除く）、鉄鋼等であり、それぞれ19.7%、33.1%、17.4%増加した。これらの中にはエンジン部品や自動車部品、完成車等が含まれ、インドネシアの自動車マーケットの回復を反映したものとされる。

インドネシア投資省・投資調整庁（BKPM）によると、対インドネシア直接投資額（実行ベース）は、35億6,300万ドル（前年比57.4%増）であり、業種別では、自動車・輸送機器が11億3,100万ドルで、投資全体の31.7%を占めた。電気・ガス・水道が7億6,800万

¹¹¹ World Economic Forum（2023）「第18回グローバルリスク報告書2023年版」。

¹¹² 外務省（2023）「外交青書」, pp.68-69.

¹¹³ JETRO（2023）「インドネシアの貿易と環境」, 2023年11月15日。
〈<https://www.jetro.go.jp/world/asia/idn/gtir.html>〉

¹¹⁴ 前掲：JETRO（2023）「インドネシアの貿易と環境」

ドル、化学・医薬品が2億7,900万ドルだった。2023年1～6月までの日本からの新規投資の中には、ダイキン工業の新工場の設立等も含まれる。

表 32 インドネシアへの日本企業の進出状況

| 項目 | 主な内容 |
|----------|---|
| 日系企業の進出数 | 2,103社（2022年10月現在） ¹¹⁵ |
| 進出状況 | <ul style="list-style-type: none"> 自動車・輸送機器は、対インドネシア直接投資額の約30%を占める（トヨタ自動車、ダイハツ工業、三菱自動車、ホンダ技研、スズキ、ヤマハ、パナソニック、シャープ、エプソン、旭硝子等が進出） |
| 在留邦人数 | 9,712人（2023年10月現在） ¹¹⁶ |

(3) 日本との関係

2023年12月16日、閣僚共同声明で、日本・インドネシア経済連携協定（JIEPA）改正交渉が大筋合意に達したと発表した。また、この閣僚共同声明を日ASEAN特別首脳会議に合わせ開催された日・インドネシア首脳会談で、岸田文雄首相とインドネシアのジョコ・ウィドド大統領が確認した。改正交渉により、日本からインドネシアへの市場アクセスについて、新たに自動車や鉄鋼・鉄鋼製品の計19品目の関税が撤廃・引き下げられる。農産品では、米粉の関税が新たに撤廃されるほか、日本産短粒種米の低関税輸入枠（割当数量：8,500トン、枠内税率は1キログラム当たり450ルピア（約4.1円、1ルピア＝約0.009円））の設定や、サービス貿易の倉庫業・不動産業の市場アクセス改善などが実現する。また、鉄鋼などの特定用途免税制度（USDFS）¹¹⁷の改善なども図られる¹¹⁸。

(4) 当該国の製造業の状況

BKPMによると、2022年の対内直接投資額（実行ベース）は456億500万ドルで、前年から46.7%増加して過去最高を更新した。国・地域別の投資は、シンガポールからは基礎金属・金属製品・非機械および器具、鉱業分野への投資であり、中国からはニッケル精錬所、高速鉄道プロジェクト関連、寧徳時代新能源科技（CATL）の子会社・寧波普勤時代（CBL）のニッケルの採掘・製錬を含むEV用バッテリーの統合事業等への投資案件がある。主要な投資案件とみられる。韓国からは、ロッテ・ケミカル・タイタン・ホールディングス（LCT）の大規模な石油化学コンビナート事業等がある。

インドネシアにおけるニッケル精錬所等をはじめ基礎金属・金属製品・非機械および器

¹¹⁵ 前掲：外務省（2022）「海外進出日系企業拠点数調査」

¹¹⁶ 前掲：外務省（2023）「海外在留邦人数調査統計」

¹¹⁷ USDFSは、自動車・同部品、スチールロール、電気・電子部品、鉱業、採掘、建設用建機などを対象に、15%前後の関税率を0%に低減できる制度。利用に当たっては1年間の輸入計画の提出や工業検査証明書取得など、厳格な運用管理が求められている。

¹¹⁸ 外務省（2023）「日・インドネシア経済連携協定改正議定書大筋合意の概要」, 2023年12月. <<https://www.mofa.go.jp/mofaj/files/100595762.pdf>>

具への投資は全体の約 2 割を占める。EV バッテリー産業では、前述の CATL、現代自動車、LG エナジーソリューションなど、中国・韓国系企業の大型投資が進展している。世界最大のニッケル産出量を有するインドネシアは、豊富なニッケルを用いた EV 用バッテリーの国産化を進めている¹¹⁹。

(5) 地理的・地政学的重要性

① 当該国と日本との友好関係

政治関係については、日本とインドネシアは 60 年以上にわたり伝統的な友好国であると同時に、民主主義等の基本的価値や法の支配に基づく自由で開かれた国際秩序の構築という目標を共有する戦略的パートナーとして、二国間協力に加えて、地域や国際社会の諸課題について連携している。2018 年には日・インドネシア外交関係開設 60 周年を迎え、各種の記念事業が実施された。

2022 年 4 月には日・インドネシア首脳会談において、両首脳は、地域が多くの課題に直面する中、両国が戦略的パートナーとして、多くの本質的な原則を共有する「自由で開かれたインド太平洋 (FOIP)」と「インド太平洋に関する ASEAN アウトルック (AOIP)」の実現に向けて協力を強化することで一致した。

② 経済安全保障やその他リスクに関する状況

経済成長および投資誘致を外交戦略の中心に据えるジョコ政権は、米中対立の深化（日米のインド太平洋戦略と中国の一帶一路構想との間の競争関係）を契機と位置づけ、安全保障上のリスクを最低限に抑えようとしている¹²⁰。

地域大国でもあるインドネシアとの協力は、地域の平和や安定にとっても重要であり、安全保障・防衛協力分野では、2015 年 12 月に日本と ASEAN 加盟国との間で初となる外務・防衛閣僚会合を開催し、2021 年 3 月には第 2 回会合を開催した。第 2 回会合に際しては、防衛装備品・技術移転協定を締結した。

インドネシアにおけるリスクについて、世界経済フォーラムの「エグゼクティブ・オピニオン調査 (EOS)」は、主要経済国の累積債務危機、国家間の紛争、急激な物価変動・長期化するインフレ、デジタル・サービスの供給不足とデジタル格差 (不平等)、戦略資源の地政学的な (覇権) 争い等を挙げた¹²¹。

¹¹⁹ 前掲：JETRO (2023) 「インドネシアの貿易と環境」。

¹²⁰ 相澤伸広 (2021) 「インドネシアの安全保障観」, SSDP 安全保障・外交政策研究会, 2021 年 6 月。〈<http://ssdpaki.la.coocan.jp/proposals/77.html>〉

¹²¹ World Economic Forum (2023) 「第 18 回グローバルリスク報告書 2023 年版」。

4.3.7. マレーシア

(1) 当該国の概況

2021年8月に発足したイスマイル・サブリ政権は、内政の安定化を図りつつ、ポスト・コロナの経済回復に注力してきたが、2022年11月に総選挙が行われた結果、アンワル新政権が発足した¹²²。総選挙では、82議席を獲得した政党連合・希望連盟（PH）を含め、主要政党連合がいずれも単独過半数の議席数に届かなかったため、前政権の与党とは異なり、国民戦線（BN）や東マレーシアの諸政党等との連立政権を構成し、事実上の政権交代が行われた¹²³。2023年12月12日には、内閣改造が行われ、経済成長や生活コスト上昇の問題に確実に効果的に取り組むため、人工知能やデジタルトランスフォーメーションへの探求と、エネルギーの移行（エネルギー移行ロードマップ）を果たすことを掲げた¹²⁴。経済状況について、2022年のマレーシアの実質GDP成長率は、COVID-19関連規制の大幅緩和や国境再開に伴い、通年で8.7%を記録し、2000年以来、22年ぶりの高水準だった¹²⁵。

(2) 日本企業の進出状況、ビジネス展開状況

日本企業の進出数は、2022年10月現在、1,856社であり、2022年の日本企業による投資案件について、製造業では、東洋インキ（ラミネート接着材の生産設備の拡充）、フェローテックホールディングス（半導体・新工場建設）等の電気・電子分野が中心である。それ以外では、NTTによる第6データセンターの着工やNXマレーシア（NIPPON EXPRESS HD傘下）によるクアラルンプール国際空港内の新倉庫開設等である¹²⁶。進出している日本企業は、電気・電子製品が最も多く、製造業の進出企業の30%超を占めている¹²⁷。日系企業の進出先は、セランゴール州、ペナン州、ジョホール州に拠点が集中し、日立、パナソニック、富士通、三菱電機、ソニー、ルネサス、シャープ、富士電機など大手が進出している¹²⁸。

マレーシアからの対日輸出を品目別にみると、天然ガスが全体の3割超を占める（日本はマレーシアの液化天然ガスの仕向け先の4割超を占める）。それ以外の品目は、集積回路、石油および歴青油、映像機器、通信機器が続く。一方で、日本からの輸入では、集積回路

¹²² 外務省（2023）「外交青書」, pp.72-73.

¹²³ JETRO（2022）「第10代首相に希望連盟アンワル氏、事実上の政権交代へ」, 2022年11月25日. <<https://www.jetro.go.jp/biznews/2022/11/7afaa0ee15c96d39.html>>

¹²⁴ JETRO（2023）「アンワル政権2年目に向け内閣改造、閣僚増員で体制強化へ」, 2023年12月14日. <<https://www.jetro.go.jp/biznews/2023/12/a548514a6725b548.html>>

¹²⁵ JETRO（2023）「マレーシアの貿易と投資」, 2023年10月10日. <<https://www.jetro.go.jp/world/asia/my/gtir.html>>

¹²⁶ 前掲：JETRO（2023）「マレーシアの貿易と投資」

¹²⁷ マレーシア日系企業リスト（業種別一覧）. <<https://connection.com.my/industrycategory/>>

¹²⁸ 国立研究開発法人科学技術振興機構 研究開発戦略センター（2022）「戦略提案・報告書 ASEAN 諸国の科学技術情勢～マレーシア～」.

<<https://www.jst.go.jp/crds/pdf/2022/OR/CRDS-FY2022-OR-02.pdf>>

が全体の 2 割を占め、次いで石油および同製品、半導体デバイス、自動車部品、乗用車が続く¹²⁹。

表 33 マレーシアへの日本企業の進出状況

| 項目 | 主な内容 |
|----------|--|
| 日系企業の進出数 | 1,856 社 (2022 年 10 月現在) ¹³⁰ |
| 進出状況 | <ul style="list-style-type: none"> 進出企業のうち、電気・電子部品企業（半導体関連）が約 30% 超を占める 近年の対マレーシアの直接投資企業も、電気・電子製品が中心である（東洋インキ、フェローテックホールディングス） |
| 在留邦人数 | 20,657 人 (2023 年 10 月現在) ¹³¹ |

(3) 日本との関係

1982 年にマハティール首相（当時）が提唱した「東方政策」、頻繁な要人往来、直接投資や貿易・技術協力などを通じた緊密な経済関係、活発な文化・留学生交流に支えられ、二国間関係は全般的に良好であった。2015 年には、ナジブ首相（当時）訪日の際に、地域や国際社会の幅広い課題について、今後一層協力を強化する「戦略的パートナーシップについての日マレーシア共同声明」が発出された。2022 年にはマハティール首相が 1982 年に開始した日・マレーシア間の友好関係の基盤である東方政策 40 周年を迎えた。

2023 年には、アンワル首相訪日に際して日マレーシア共同声明を発出し、両国関係を「包括的・戦略的パートナーシップ」に格上げした¹³²。

(4) 当該国の製造業の状況

製造業（電気機器）が GDP の 24% を占める¹³³。業種分野は、エレクトロニクスの他、石油・ゴム・プラスチック等。エレクトロニクスでは特に半導体産業が強く、多国籍企業の米国・インテル、ドイツ・インフィニオン・テクノロジーズ、オランダ・Nexperia、オーストリア・AT&S などが、パワー半導体を含む集積回路や周辺材料、部品などの設備投資を行い、海外オフショア型開発・生産を通して、マレーシアは半導体の重要なサプライチェーンとなっている。今後成長が期待される分野は、エレクトロニクス産業、航空・宇宙産業、医療機器、機械加工・機械装置、化学産業、バイオ産業等の強化が見込まれている。一方で、マレーシア製造業が直面している課題として、研究・開発のための資金不足が挙

¹²⁹ 前掲：JETRO（2023）「マレーシアの貿易と投資」

¹³⁰ 前掲：外務省（2022）「海外進出日系企業拠点数調査」

¹³¹ 前掲：外務省（2023）「海外在留邦人数調査統計」

¹³² 外務省「国・地域」情報：マレーシア

〈<https://www.mofa.go.jp/mofaj/area/malaysia/data.html>〉

¹³³ Malaysian Economic Statistics Review (MESR) Vol 2 2023.

〈https://v1.dosm.gov.my/v1/index.php?r=column/cthemedByCat&cat=481&bul_id=eDdzbE1OVGt_xbzlMz2pWZkVRTkZtdz09&menu_id=azJjRWpYL0VBYU90TVhpclByWjdMQT09〉

げられる¹³⁴。

2022年のマレーシアへの対内直接投資（国際収支ベース、ネット、フロー）は、前年を上回り、過去最高（746億リンギ）を記録した。製造業への投資が大きく増加し、全体の66.4%を占める。製造業においては、北部ペナン州を中心に広がる電気電子分野の裾野の広さ、高度人材も一定程度獲得可能であること、英語力の高さといった従来の強みに加え、米中間の緊張関係の高まりを背景としたリスク回避行動も、特に高付加価値半導体などでマレーシアが投資先として選定される要因と考えられている。近年の投資計画の多くは、太陽光電池や電気自動車（EV）用電池など環境分野が多い¹³⁵。

（5）地理的・地政学的重要性

①当該国と日本との友好関係

2023年12月16日に、二国間首脳会談で「包括的・戦略的パートナーシップに関する日・マレーシア共同声明」が公表された。同声明では、ASEAN中心性・一体性を尊重しつつ、自由で開かれたインド太平洋と、インド太平洋に関するASEANアウトックとの相乗効果を追求していくことを示した。経済環境では、経済産業省とマレーシア政府との間の航空機産業協力に係る協力覚書に基づく航空機分野における協力の進展や、災害管理のための情報通信技術、オープンで安全かつ強靱な5Gネットワークと将来のための革新的なネットワーク技術の開発と採用、サイバーセキュリティ、IoT、スマートシティ、クラウド・コンピューティング、デジタル経済、スマート製造業を含む放送及びICT分野における協力を促進することを確認した。あわせて、地域内外において、透明で、多様で、安全で、持続可能な、信頼できるサプライチェーンの重要性を認識しつつ、サプライチェーンの強靱性の強化を含む経済安全保障に関して協力する意図を共有した¹³⁶。

②経済安全保障やその他リスクに関する状況

マレーシアは、マレー半島の「半島マレーシア」とボルネオ島の「東マレーシア」から成る、インド洋と太平洋の結節点に位置し、南シナ海とマラッカ海峡に面した地政学的に重要な国である。

ASEAN域内では、結束を揺るがすミャンマー情勢、域外からは、南シナ海の実効支配を強める中国に対する緊張関係、米中間の東南アジア各国をめぐる駆け引きという圧力が存在。米中は、短期的に直接的な対立よりも、特に東南アジア・南アジア・大洋州地域諸国を自らの陣営に取り込む競争を繰り広げるといった見方がある¹³⁷。なお、マレーシアとは、

¹³⁴ 前掲：国立研究開発法人科学技術振興機構 研究開発戦略センター（2022）「戦略提案・報告書 ASEAN 諸国の科学技術情勢～マレーシア～」

¹³⁵ 前掲：JETRO（2023）「マレーシアの貿易と投資」

¹³⁶ 外務省（2023）「包括的・戦略的パートナーシップに関する日・マレーシア共同声明」，2023年12月16日．〈<https://www.mofa.go.jp/mofaj/files/100596111.pdf>〉

¹³⁷ PWC（2023）「2023年主要地域における地政学リスク動向」

〈<https://www.pwc.com/jp/ja/knowledge/column/geopolitical-risk-column/vol1.html>〉

「日・マレーシア防衛装備品・技術移転協定」に署名している¹³⁸。

また、マレーシアにおけるリスクについて、世界経済フォーラムの「エグゼクティブ・オピニオン調査 (EOS)」は、生活費危機、急激な物価変動・長期化するインフレ、主要経済国の累積債務危機、地経学上の対立、人為的な環境被害や災害、著しい資源・物品供給の危機、極端なコモディティショックや相場の乱高下、長期化する経済停滞、サイバーセキュリティ対策の失敗等を挙げた¹³⁹。

4.3.8. ベトナム

(1) 当該国の概況

ベトナムは、ベトナム共産党が一党支配する統治体制となっており、中央委員候補者のうち、政治局での序列の高い人物が党書記長、国家主席（大統領）、政府首相、国会議長の四つの要職に就任する。ベトナムでは、共産党の最高指導者である書記長、政府の長である首相、国家元首である国家主席の 3 人の首脳によるトロイカ体制によって政治運営が行われる。第 13 期（2021～2025 年期）の党書記長は、グエン・フー・チョン (Nguyen Phu Trong) が再任され、国家主席はボー・ティ・アイン・スアン (Vo Thi Anh Xuan) が暫定的に国家主席代行を務めている（国家主席であったグエン・スアン・フック氏は COVID-19 パンデミック対応、政府高官の汚職事件等で辞任）、首相はファム・ミン・チン (Pham Minh Chinh) 氏である¹⁴⁰。

ベトナム全体は、東南アジア第 3 位の人口を有し、中間所得層が急増していることから、有望な市場でもある。現在、インフレ抑制などのマクロ経済安定化、インフラ整備や投資環境改善を通じた外資誘致を通じ、安定的な経済成長の実現に取り組んでいる¹⁴¹。2022 年、ベトナムは COVID-19 パンデミックからの反動もあり、実質 GDP 成長率は 8.0% で、力強い経済回復を見せた。

(2) 日本企業の進出状況、ビジネス展開状況

日本企業の進出数は、2022 年 10 月現在、2,373 社であり、キヤノン、パナソニック、ホンダ、トヨタ、富士通、日本電産、ブリヂストン、富士ゼロックス、マブチモーター、イオン、ファミリーマート、ルネサンス等の企業が進出している。

2022 年の日本からの直接投資（認可ベース、出資・株式購入を除く）は、新規・拡張の合計で 406 件、47 億 9,300 万ドルである。日本からの投資案件は、丸紅と東京ガスが出資

¹³⁸ 外務省（2018）「日・マレーシア防衛装備品・技術移転協定」, 2018 年 4 月 18 日。
〈https://www.mofa.go.jp/mofaj/press/release/press4_005926.html〉

¹³⁹ World Economic Forum (2023)「第 18 回グローバルリスク報告書 2023 年版」。

¹⁴⁰ 国際協力銀行（2023）「ベトナムの投資環境」。

〈https://www.jbic.go.jp/ja/information/investment/image/inv_vietnam02.pdf〉

¹⁴¹ 外務省「国・地域」情報：ベトナム

〈<https://www.mofa.go.jp/mofaj/area/vietnam/data.html>〉

する北部クアンニン省でのLNG火力発電所案件（20億ドル）が最大で、日本の新規投資認可額の6割近くを占める。また、FUJI SOLAR社が北部フート省・太陽電池研究・製造案件（2億ドル）に出資している。製造業の投資件数は、拡張（100件）が新規（41件）を大きく上回った。既に進出している企業が、生産の強化やビジネスの拡大を図る傾向が強く表れた。一方、ベトナムにおけるビジネスリスク（国家主席の辞任等の政治の混乱、経済成長の低迷、政府高官の汚職等）が顕在化し、政府内で要職の空席に加え、処分を恐れて判断を避ける傾向が強まり、行政における各種許認可や手続きの遅延が発生した。

また、日本企業の進出・投資にあたっては、問題点として、①従業員の賃金上昇、②調達コストの上昇、③為替変動、④通関等諸手続きが煩雑、⑤税務（法人税、移転価格課税など）の負担等が挙げられている¹⁴²。

表 34 ベトナムの日本企業の進出状況

| 項目 | 主な内容 |
|----------|---|
| 日系企業の進出数 | 2,373社（2022年10月現在） ¹⁴³ |
| 進出状況 | <ul style="list-style-type: none"> 進出企業の生産強化やビジネス拡大が活発化（キヤノン、パナソニック、ホンダ、トヨタ、富士通、日本電産、ブリヂストン、富士ゼロックス、マブチモーター、イオン、ファミリーマート、ルネサンス等） |
| 在留邦人数 | 18,949人（2023年10月現在） ¹⁴⁴ |

（3）日本との関係

1973年9月21日、日越間の外交関係が樹立。1978年末のベトナム軍カンボジア侵攻に伴い、対越経済協力を見合せたが、1991年10月のカンボジア和平合意を受け、翌年455億円を限度とする円借款供与。現在、日越関係は「アジアにおける平和と繁栄のための広範な戦略的パートナーシップ」の下、政治、経済、安全保障、文化・人的交流など幅広い分野で緊密に連携している。

日本とベトナムは「アジアにおける平和と繁栄のための広範な戦略的パートナーシップ」の下で、様々な分野で協力を進展させることを発表した。具体的には、①ベトナム軍へのサイバーセキュリティ能力構築支援といった安全保障分野における協力の推進、②サプライチェーン多元化、デジタルトランスフォーメーション、技術革新を含む経済分野の協力、③技能実習生送出に関するプラットフォーム構築、④防災能力向上のための地球観測衛星打ち上げ支援などにおける協力の推進について確認した。2023年12月には、日越首脳会談の結果、「アジアと世界における平和と繁栄のための包括的戦略的パートナーシップ」に格上げし、経済的連携面では、日本が独立かつ自立したベトナム経済の発展を引き続き支

¹⁴² JETRO（2023）「ベトナムの貿易と投資」, 2023年11月24日。

〈<https://www.jetro.go.jp/world/asia/vn/gtir.html>〉

¹⁴³ 前掲：外務省（2022）「海外進出日系企業拠点数調査」

¹⁴⁴ 前掲：外務省（2023）「海外在留邦人数調査統計」

援していくと表明した¹⁴⁵。

技能実習生を中心に在日ベトナム人の数も増加しており、2011年の約4万人から2021年12月末には約43万人を超え、国別在留外国人数で中国に次いで2番目に多い数字となっている¹⁴⁶。

(4) 当該国の製造業の状況

ベトナムの製造業がGDP全体の25%近くを占め、2022年の貿易額は過去最高を記録した。実質GDP（国内総生産）成長率は2021年2.56%とコロナ禍の影響を受け2011年以降最低となったが2022年は8.02%に回復した。製造業は4.82%と堅調であった。しかし、2023年は5.05%に鈍化している¹⁴⁷。

ベトナムの2022年の対内直接投資（認可ベース、出資・株式取得を除く）は、新規・拡張の合計で3,362件（前年比15.3%増）、認可額は238億4,700万ドル（24.9%減）であった。認可件数は3年ぶりに増加し、認可額は新型コロナ禍前の2019年とほぼ同水準を保っている。業種別では、製造業が1,222件、160億1,400万ドルで、件数・金額ともに首位だった。直接投資の案件例として、大型のものは、レゴ（デンマーク）の玩具工場の新規投資（約13億2,000万ドル）、サムスン電機（韓国）の半導体パッケージ基板工場の拡張投資（9億2,000万ドル）が挙げられる¹⁴⁸。

中国以外の国や地域に生産拠点を分散させる「チャイナプラスワン」の最有力国としてベトナムが注目され、ベトナム自身も製造業の成長に向けて外資企業を積極的に誘致、多くの外国企業が進出してきた。インテル、サムスン、LG、パナソニック等の大手電子機器メーカーが生産拠点を置いたことで、ベトナムは2019年の製品輸出金額においてシンガポールに次ぐASEAN第2位となった。韓国系や日系企業等がベトナムを進出先として選択する理由として、安価な人件費、ベトナム人の性格と儒教文化圏の企業文化との親和性の高さ、自由貿易協定の締結で輸出向け製品の事業環境が良好といったことが挙げられている。一方、課題としては、部品や原材料の現地調達率の低さ、大企業による大量雇用による人材難、工業団地の不足、土地価格や賃貸料の上昇などが挙げられている¹⁴⁹。

¹⁴⁵ JETRO（2023）「トゥオン国家主席が訪日、日ベトナム関係を包括的戦略的パートナーシップに格上げ」、2023年12月7日。

〈<https://www.jetro.go.jp/biznews/2023/12/0e89c9cb5cc1f724.html>〉

¹⁴⁶ 外務省（2023）「外交青書」、p.72。

¹⁴⁷ ロイター（2023）「ベトナム、23年成長率5.05%に減速 政府目標に届かず」

〈<https://jp.reuters.com/markets/japan/funds/QK3XHLHVHBN4BBJS6GIUUDZG2Y-2023-12-29/>〉 *2024年3月14日調べ

¹⁴⁸ 前掲：JETRO（2023）「ベトナムの貿易と投資」

¹⁴⁹ 矢野経済研究所（2021）『「ベトナムにおける製造業の動向に関する調査（2021年）」調査概要』〈https://www.yano.co.jp/press-release/show/press_id/2787〉

(5) 地理的・地政学的重要性

①当該国と日本との友好関係

ベトナムは、東南アジア第3位の人口を有し、中間所得層が急増していることから、有望な市場でもある。日本とベトナムの関係は、1973年9月21日、日越間の外交関係が樹立した。1978年末のベトナム軍カンボジア侵攻に伴い、対越経済協力を見合せたが、1991年10月のカンボジア和平合意を受け、翌年455億円を限度とする円借款供与した。現在、日越関係は「アジアにおける平和と繁栄のための広範な戦略的パートナーシップ」の下、政治、経済、安全保障、文化・人的交流など幅広い分野で緊密に連携している¹⁵⁰。

②経済安全保障やその他リスクに関する状況

ベトナムは、南シナ海のシーレーンに面し、中国と長い国境線を有する地政学的に重要な国である。日本とベトナムは、「アジアにおける平和と繁栄のための広範な戦略的パートナーシップ」の下で、様々な分野で協力を進展させている。2022年5月には岸田総理大臣がベトナムを訪問し日越首脳会談を行った。同会談において両首脳は、①ベトナム軍へのサイバーセキュリティ能力構築支援といった安全保障分野における協力の推進、②サプライチェーン多元化、デジタルトランスフォーメーション、技術革新を含む経済分野の協力、③技能実習生送出に関するプラットフォーム構築、④防災能力向上のための地球観測衛星打ち上げ支援などにおける協力の推進について確認した¹⁵¹。

ベトナムにおけるリスクとして、世界経済フォーラムの「エグゼクティブ・オピニオン調査 (EOS)」は、急激な物価変動・長期化するインフレ、感染症の広がり、戦略資源の地政学的な(覇権)争い、自然災害と異常気象、サイバーセキュリティ対策の失敗等を挙げた¹⁵²。

4.3.9. タイ

(1) 当該国の概況

タイでは、2023年9月にセター・タウィーンシン首相による新政権が発足した。新政権の発足までは複雑であり、2023年5月に下院総選挙が実施され、旧野党であった「前進党」が大躍進し第1党(151議席)となり、次いで旧野党の「貢献党」(元タクシン首相派)が141議席を獲得した。「前進党」の党首のピター氏は、同年7月の首班指名選挙で否決され、貢献党は前進党との連立の枠組みを解消し、前進党を除いた旧与野党を含めた新たな連立の枠組みを発表し、セター政権が発足した。

近年は、積極的な経済外交を展開し、中東やEU諸国との関係強化の動きが見られる。国際的に関心の高い問題に対して声明を発表する等、国外の情勢や地球規模課題等について立場表明を行っている。また地政学的にメコン地域の中核を成すことから地域情勢にも

¹⁵⁰ 前掲：外務省「国・地域」情報：ベトナム

¹⁵¹ 前掲：外務省(2023)「外交青書」

¹⁵² World Economic Forum (2023)「第18回グローバルリスク報告書2023年版」。

積極的に関与する傾向が見られ、エーヤワディ・チャオプラヤ・メコン経済協力戦略会議（ACMECS）の活性化、タイ国際協力機構（TICA）を通じた周辺国への国際協力やハード・ソフト両面の連結性の向上等の取り組みが上げられる。2021年半ばから3年間、タイはASEANの対日調整国（ASEANを代表して日本との関係で窓口を務める国）を務めている¹⁵³。

タイの経済状況は、実質GDP成長率は2.6%と前年の1.5%からプラス幅が拡大した。財・サービスの輸出入の伸びが鈍化したものの、個人消費が大きく拡大した¹⁵⁴。

（2）日本企業の進出状況、ビジネス展開状況

日本企業の進出数は、2022年10月現在、5,856社であり、ASEAN最多の進出企業数である。タイは、東南アジア最大の海外自動車企業集積地であり、日本の自動車産業ではトヨタ、日産、ホンダ、マツダ、スズキ、いすゞ自動車、日野自動車、三菱自動車が進出している。その他業種では、東芝、キヤノン、パナソニック、NTT、AGC、東レ、味の素、王子製紙、ダイキン、クボタ、三井住友銀行、三菱UFJ銀行、博報堂、電通、丸紅、伊藤忠商事、商船三井、川崎汽船、全日本空輸（ANA）、イオン等が進出している¹⁵⁵。

一方で、タイでは高齢化の進行、若年労働力不足、賃金上昇等によって「低廉な労働力」がメリットの生産拠点ではなくなりつつある。過去、日本企業の投資誘致に成功してきたタイ投資委員会（BOI）の危機感を有している。

2022年の日本からの直接投資は、申請ベースで293件、508億バーツであり、国・地域別で中国に次いで2位となった。一方、認可ベースでは、先端材料・ナノ材料、ハードディスクドライブ（HDD）の製造事業等の大型案件があり、日本は認可件数（216件）、認可額（500億バーツ）でトップだった。2022年は、中国のBYD、台湾の鴻海精密工業が電気自動車（EV）の完成車工場の建設を申請するなどEV関連の投資が増加しており、これらの新たな投資の動きに合わせて、日本CMKが2022年9月、車載向けプリント基板生産の新工場の建設を発表（投資額は約250億円）、タイ東海カーボン・プロダクトが2022年9月、タイヤ補強材・カーボンブラック生産の新工場の建設を発表（投資額は約346億円）した¹⁵⁶。

電気自動車産業の育成は、BOIの優先事項であり、日本との緊密な連携が期待されている¹⁵⁷。

¹⁵³ 外務省「国・地域」情報：タイ〈<https://www.mofa.go.jp/mofaj/area/thailand/data.html>〉

¹⁵⁴ JETRO（2023）「タイの貿易と投資」、2023年10月26日。

〈<https://www.jetro.go.jp/world/asia/th/gtir.html>〉

¹⁵⁵ <https://idgjapan.com/japanese-companies-in-thailand/> *2024年3月14日調べ

¹⁵⁶ 前掲：JETRO（2023）「タイの貿易と投資」

¹⁵⁷ JETRO（2022）「タイの投資誘致政策と日本のAJIFが目指す未来」、2022年8月12日。

〈<https://www.jetro.go.jp/biz/areareports/2022/3cad64b2fee02ca3.html>〉

表 35 タイへの日本企業の進出状況

| 項目 | 主な内容 |
|----------|---|
| 日系企業の進出数 | 5,856 社 (2022 年 10 月現在) ¹⁵⁸ |
| 進出状況 | <ul style="list-style-type: none"> 東南アジア最大の自動車企業の集積地 (トヨタ、日産、ホンダ、マツダ、スズキ、いすゞ自動車、日野自動車、三菱自動車が進出) 電気自動車 (EV) 関連の投資も増加し、関連企業の工場建設等の新規投資も進む |
| 在留邦人数 | 72,308 人 (2023 年 10 月現在) ¹⁵⁹ |

(3) 日本との関係

日タイ両国は 600 年にわたる交流・友好の歴史。両国の皇室・王室間の親密な関係を基礎に、政治、経済、文化等で緊密且つ重層的な関係を構築してきた。

2022 年は日・タイ修好 135 周年の年であり、両国の「戦略的パートナーシップ」(2012 年)以降、二国間関係を深化させ、2022 年 5 月には日タイ首相会談を踏まえ、「日・タイ防衛装備品・技術移転協定」に署名した。同年 11 月には、今後 5 か年の経済分野の協力の方向性を定めた「日タイ戦略的経済連携 5 か年計画」に署名するとともに、「包括的戦略的パートナーシップ」に格上げした¹⁶⁰。「日タイ戦略的経済連携 5 か年計画」では、タイ政府が推進する「バイオ・循環型・グリーン (BCG) 経済」モデル、日本の「アジア・ゼロエミッション共同体構想」などを通じて、最新のエネルギー技術開発や低炭素エネルギー事業への投資で協力を深める方針を確認した¹⁶¹。

(4) 当該国の製造業の状況

タイ投資委員会 (BOI) によると、2022 年の外国企業による対内直接投資 (FDI) 申請額は、4,340 億バーツ (前年比 36.3%増) となった。国・地域別では中国が 774 億バーツ (前年比 2.1 倍) で首位であり、次いで、日本が 508 億バーツ (前年比 36.6%減)、米国が 503 億バーツ (前年比 72.5%増)、台湾が 452 億バーツ (前年比 2.1 倍)、シンガポールが 443 億バーツ (前年比 49.5%増) と続いた。投資額の多くは、タイの主要な工業地域である東部経済回廊 (EEC) への FDI 投資額が 2,553 億バーツ (前年比 74.2%増) と最も多く、全体の 58.8%が同地域に集中した。中国からの投資拡大の背景は、2024 年から本格的に始まる EV 生産であり、タイでは 2024 年末までの国内生産開始を条件とした EV 補助金政策を実施している。このため、中国メーカーが一斉に EV の生産を開始する (比亞迪汽車 (BYD) が年間 15 万台、長安汽車が 10 万台、長城汽車 (GWM) が 8 万台)。また、台湾のフォックスコンが 5 万台を生産する計画とされ、これらの生産能力を単純に合計すると 40 万台を超

¹⁵⁸ 外務省 (2022) 「海外進出日系企業拠点数調査」

¹⁵⁹ 外務省 (2023) 「海外在留邦人数調査統計」

¹⁶⁰ 外務省 (2023) 「外交青書」, p.70.

¹⁶¹ JETRO (2022) 「日タイ首脳会談、包括的戦略的パートナーシップへと関係を格上げ」, 2022 年 11 月 28 日. <<https://www.jetro.go.jp/biznews/2022/11/d9a22cd4c381aa44.html>>

える（タイの年間自動車生産台数の約 20%を占める）。タイでは、「タイランド 4.0」（国家指針）や、「バイオ・循環型・グリーン（BCG）経済モデル」（国家戦略）の下、EV、デジタル、スマート電子機器、BCG 経済、クリエイティブ産業などの重要分野にて、他産業よりも手厚い投資優遇措置（法人税の減免など）を提供している¹⁶²。

（5）地理的・地政学的重要性

①当該国と日本との友好関係

日タイ両国は 600 年にわたる交流の歴史を持ち、伝統的に友好関係を維持している。長年の両国の皇室・王室間の親密な関係を基礎に、政治、経済、文化等幅広い面で緊密且つ重層的な関係を築いており、人的交流は極めて活発である。2022 年は日タイ修好 135 周年を迎えた。日タイの経済関係は、前述の「包括的戦略的パートナーシップ」、「日タイ戦略的経済連携 5 か年計画」の通り、中長期にわたり、継続的な進化・拡大していく方向であり、2022 年から 2026 年までの 5 年間で①人材育成、規制改革、イノベーション、②バイオ・循環型・グリーン（BCG）経済、③インフラの協力を進めることになっている¹⁶³。

②経済安全保障やその他リスクに関する状況

タイは、1967 年の「バンコク宣言」により誕生した ASEAN の原加盟国の一つであり、また、メコン地域の中心に位置し、地政学的に重要な国である。ASEAN 諸国との連携と、日本、米国、中国といった主要国との協調を外交の基本方針である。タイは、ASEAN 議長国として、「インド太平洋に関する ASEAN アウトルック(AOIP)」を策定した。これは、ASEAN の中心性を堅持し、多極アジアを実現することを掲げている¹⁶⁴。

タイにおけるリスクについて、世界経済フォーラムの「エグゼクティブ・オピニオン調査 (EOS)」は、主要経済国の累積債務危機、生活費危機、人為的な環境被害や災害、デジタル・サービスの供給不足とデジタル格差（不平等）、地経学上の対立等を挙げた¹⁶⁵。

4.3.10. サウジアラビア

（1）当該国の概況

サウジアラビアは、君主制を敷いており、国家元首はサルマン・ビン・アブドゥルアジーズ・アール・サ우드国王（以下、サルマン国王）である。議会は、立法権はないが、

¹⁶² 前掲：JETRO（2023）「タイの貿易と投資」

¹⁶³ 外務省（2022）「包括的戦略的パートナーシップへの関係格上げに関する タイ王国政府と日本国政府との間の共同声明」, 2022 年 11 月 17 日。

¹⁶⁴ 菊池努（2019）「「インド太平洋」戦略の地政学—地域秩序の主体は誰か—」, 国際問題, No.687, 2019 年 12 月. <https://www2.jiia.or.jp/kokusaimondai_archive/2010/2019-12_001.pdf?noprint>

¹⁶⁵ World Economic Forum（2023）「第 18 回グローバルリスク報告書 2023 年版」.

諮問評議会（マジュリス・アール・シューラー）がある¹⁶⁶。

2017年4月、サルマン国王及びムハンマド副皇太子（当時）が主導して「サウジ・ビジョン2030」を策定し、2030年までのサウジ社会の石油依存からの脱却と産業多角化等を中心とした改革の実現に向けて各種政策を進めている¹⁶⁷。

2022年のサウジアラビアの実質GDP成長率は、原油価格の高騰や民間企業の投資が非石油部門の成長を促し過去11年間で最高の8.7%となった。2022年の産業別の成長率は、鉱業・採石が16.0%、運輸・倉庫・通信が9.1%、製造業が7.9%と続いた。また、サウジアラビア人の失業率は、2021年までは2桁台で推移していたが、2022年は9.4%となった。¹⁶⁸

（2）日本企業の進出状況、ビジネス展開状況

日本企業の進出数は、2022年10月現在、110社であり、クボタ（鋳鋼製品）、ジェイ・パワーシステムズー丸紅メタル（電力海底ケーブル製造・販売）、東洋紡ー伊藤忠商事、東レ（海水淡水化向けRO膜エレメントの製造・販売）、いすゞ自動車（トラック組み立て・販売）、アズビル（調整弁の製造・販売）、東邦チタニウム（スポンジチタン製造）、SPDC [シャルク]（ポリエチレン、エチレングリコール）、住友化学 [ペトロラービグ]（エチレン、ポリエチレン）、三菱化学 [サマック]（メタクリル酸メチル、アクリル樹脂）、MHPS（ガスタービン補修工場）、日立製作所（ガス絶縁開閉装置の設計、建設）、日本郵船（キングアブドラ港運営）、荏原製作所（ポンプの製造販売）等の企業が進出し投資している。

投資の背景には、後述する「サウジアラビア20230」に基づく、「日・サウジ・ビジョン2030」による戦略的パートナーシップの締結が挙げられる。同ビジョンは、新しい日本ーサウジアラビアの協力の羅針盤として、脱石油依存と雇用創出のためサウジが追求する「サウジビジョン2030」と、GDP600兆円の達成に向けて日本が追求する「日本の成長戦略」のシナジーを目指すことを掲げ、日本の経済セクターでは、全面バックアップの下、36の企業間の覚書を締結している¹⁶⁹。

日本からサウジアラビア向けの輸出は、自動車を中心とする輸送用機器が約70%を占める。日本のサウジアラビアからの主要輸入品目は、鉱物性燃料（原油）が大半を占めている。

¹⁶⁶ 諮問評議会は、国王に意見具申を行う機能を有している。定員は150名であり、任期4年である。2013年1月には、女性枠（最低20%）が設定され、30名の女性議員が任命された。

¹⁶⁷ 外務省「国・地域」情報：サウジアラビア王国
〈<https://www.mofa.go.jp/mofaj/area/saudi/data.html>〉

¹⁶⁸ JETRO（2023）「サウジアラビアの貿易と投資」, 2023年9月26日。
〈https://www.jetro.go.jp/world/middle_east/sa/gtir.html〉

¹⁶⁹ 上田悠紀子（2021）「中東経済の動向と日本企業の進出概要」, 経済産業省中東アフリカ課
〈https://www.eduport.mext.go.jp/epsite/wp-content/uploads/2021/03/middle-east2_shiryo1.pdf〉

表 36 サウジアラビアへの日本企業の進出状況

| 項目 | 主な内容 |
|----------|--|
| 日系企業の進出数 | 110社（2022年10月現在） ¹⁷⁰ |
| 進出状況 | <ul style="list-style-type: none"> 日・サウジ・ビジョン 2030 を背景に、クボタ、ジェイ・パワーシステムズー丸紅メタル、東洋紡ー伊藤忠商事、東レ、いすゞ自動車、アズビル、東邦チタニウム、SPDC [シャルク]、住友化学 [ペトロラービグ]、三菱化学 [サマック]、MHPS、日立製作所、日本郵船、荏原製作所等が投資している。 |
| 在留邦人数 | 628人（2023年10月現在） ¹⁷¹ |

(3) 日本との関係

1955年に外交関係樹立し、日本の皇室とサウジアラビアの王室との関係は良好である。2016年9月、ムハンマド副皇太子が訪日し、「サウジ・ビジョン 2030」の下でサウジアラビアが進める改革努力を支援する包括的パートナーシップ実現への協力で一致し、2017年3月に協力の枠組みを定めた「日・サウジ・ビジョン 2030」の実施に合意した。2022年7月には、岸田総理がサウジアラビアを訪問し、ムハンマド皇太子兼首相と「日・サウジ・ビジョン 2030」の下で、脱炭素分野を含め一層幅広い協力を進めることで一致した¹⁷²。当該ビジョンは、多様性、革新性、ソフトな価値の3本柱を掲げ、農業・食料安全保障、エンタメ・メディア、医療・保健、質の高いインフラ、投資・ファイナンス、エネルギー、競争力のある産業、中小企業の能力開発、文化・スポーツ・教育の9分野に対して、57の協力プロジェクトと36の企業間覚書が締結されている。ビジネス促進のための措置として、規制緩和、インセンティブ、組織的支援の3つのビジネス促進措置（ESZ）の候補地を選定する¹⁷³。

(4) 当該国の製造業の状況

サウジアラビアの2022年の対内直接投資額（国際収支ベース、ネット、フロー）は、前年比59.1%減の79億ドル、対内直接投資額の残高（ストック）は2,689億ドルであったものの、サウジアラビア投資省（MISA）が2022年に外国企業に供与した投資ライセンス数は1万6,452件と、前年の4,477件から約3.7倍と大幅に増加した。業種別では、製造業が最多の9,179件で、情報通信の1,036件が続く。投資ライセンスが急増した最大の要因は、2020年8月に成立した反隠匿法（Anti-Concealment law）に基づく新規ライセンス取得であり、従来、海外企業（親会社）は現地法人を持たずに、パートナー企業（地場企業）が外国人労働者のスポンサーとなり、パートナー企業所属で親会社の業務を担ってきたが、パ

¹⁷⁰ 外務省（2022）「海外進出日系企業拠点数調査」

¹⁷¹ 外務省（2023）「海外在留邦人数調査統計」

¹⁷² 前掲：外務省「国・地域」情報：サウジアラビア王国

¹⁷³ 前掲：上田悠紀子（2021）「中東経済の動向と日本企業の進出概要」

一トナー企業所属で親会社の業務活動を行うことができなくなり、新規法人設立が義務付けられた影響である¹⁷⁴。

2016年に公報された「サウジ・ビジョン 2030」は、ムハンマド・ビン・サルマン（MbS）副皇太子が主導し、2030年までの今後15年間で達成すべき目標とそのための政策アジェンダをまとめたもので、①脱石油依存経済、②雇用の創出、③効率的な行政の3つの目標を達成することで、サウジアラビアが抱える課題を包括的に解決しようとするものである。脱石油依存経済では、非石油収入の拡大と、非石油輸出割合の増加を掲げた。

（5）地理的・地政学的重要性

①当該国と日本との友好関係

サウジアラビアとは、1938年に最初の外交接触があり、1955年に国交を樹立した。1971年には故ファイサル国王が国賓として来日され、故昭和天皇と会見し、1981年に皇太子・同妃両殿下（現天皇・皇后両陛下）が日本の皇室として初めてサウジアラビアを訪問した。1994年には皇太子・雅子妃両殿下がサウジアラビアを訪問され、天皇家二世代にわたる友好関係にある。

1957年、日本企業のアラビア石油株式会社がサウジアラビア政府より石油採掘権を供与され、クウェートとの旧中立地帯沖合で石油の採掘に成功し、同社の石油採掘権が失効した2000年まで継続し、両国の経済交流の基礎となっている。2017年には、前述の日本企業のサウジ進出を促す経済特区新設や、エネルギーやインフラなど幅広い分野での協力を定めた「日・サウジ・ビジョン 2030」が合意した¹⁷⁵。

②経済安全保障やその他リスクに関する状況

日本にとって最大の原油供給国であり、日本は輸入原油の約4割をサウジアラビアから調達している。サウジアラビアは、世界最大級の石油埋蔵量、生産量及び輸出量を誇るエネルギー大国であり、輸出総額の約9割、財政収入の約8割を石油に依存している。OPEC（石油輸出国機構）の指導国として国際原油市場に強い影響力を有する。若年層への雇用機会の増大、石油依存からの脱却が最重要課題であり、サウジ人労働力の積極的利用（サウダイゼーション）や石油部門以外の部門の発展に力を注いでいる。また、人材育成、民営化、外資導入、市場開放等諸改革に努めている。2016年4月、石油依存からの脱却と産業多角化を目指す「サウジ・ビジョン 2030」が発表され、観光振興や女性の労働参画推進等を含む包括的な社会経済改革の方針が示された¹⁷⁶。

サウジアラビアにおけるリスクについて、世界経済フォーラムの「エグゼクティブ・オピニオン調査（EOS）」は、生活費危機、国家間の紛争、急激な物価変動・長期化するインフレ、極端なコモディティショックや相場の乱高下、サイバー攻撃に因る重要な情報イン

¹⁷⁴ 前掲：JETRO（2023）「サウジアラビアの貿易と投資」

¹⁷⁵ 日本サウディアラビア協会「日サ関係の歩み」〈<http://www.js-society.com/sajp1.html>〉

¹⁷⁶ 前掲：外務省「国・地域」情報：サウジアラビア王国

フラとネットワークの機能停止、サイバーセキュリティ対策の失敗等を挙げた¹⁷⁷。

4.3.11. インド

(1) 当該国の概況

インドでは、2004年の第14回下院議員総選挙に引き続き、2009年の第15回下院議員総選挙でも、与党 कांग्रेस党を第一党とする統一進歩連合（UPA）が過半数を確保し、第2次 UPA 政権が発足した。

2014年4月から5月に行われた第16回下院議員総選挙では、インド人民党（BJP）が単独過半数を超えて大勝し、インド人民党（BJP）政権（ナレンドラ・モディ首相）が発足した。2019年4月から5月に行われた第17回下院議員総選挙でも、インド人民党（BJP）が単独過半数を超えて大勝し、インド人民党（BJP）政権（ナレンドラ・モディ首相）が継続している¹⁷⁸。

2022年度（2022年4月～2023年3月）の実質 GDP 成長率は7.2%と、新型コロナウイルス禍後にV字回復を遂げた前年度の9.1%をやや下回ったが、堅調な伸びを記録した。GDPのうち約6割を占める民間最終消費支出の伸びが7.5%と高い水準を維持したほか、国内総固定資本形成も11.4%と前年度に引き続き2桁成長となり、両者がインド経済をけん引した¹⁷⁹。

(2) 日本企業の進出状況、ビジネス展開状況

日本企業の進出数は、2022年10月現在、4,901社である。進出企業社数の割合（業種別）は、「製造業」が全体の48.7%と半数近くを占め、内訳は輸送用機械器具（10.4%）、電気機械器具（5.4%）、化学工業（5.4%）、金属製品（4.3%）である。製造業以外では、商社を中心とする「卸売業」が13.3%、「情報通信業」が5.8%、「運輸業・郵便業」が4.8%である¹⁸⁰。

2022年の日本からの直接投資（実行ベース）は前年比9.0%増の20億645万ドルとなり、投資国としては6位である。

製造業では、凸版印刷（現 TOPPAN）がマックス・スペシャリティー・フィルムズ（フィルムメーカー）を、クボタがエスコーツ（トラクターメーカー）を、それぞれ連結子会社化した（2022年）。また、パナソニックは配線器具の新工場を稼働開始し、CKDは空気圧機器や流体制御機器等の生産工場の建設を、三菱電機はFA制御システム製品の新工場の建設を明らかにした。また、東海理化はスマートフォンと連動する次世代型の自動車用ス

¹⁷⁷ World Economic Forum（2023）「第18回グローバルリスク報告書2023年版」。

¹⁷⁸ 外務省「国・地域」情報：インド共和国

¹⁷⁹ JETRO（2023）「インドの貿易と投資」, 2023年11月29日。

〈<https://www.jetro.go.jp/world/asia/in/gtir.html>〉

¹⁸⁰ JETRO（2022）「進出日系企業数が初めて減少（インド）」, 2022年6月28日

〈<https://www.jetro.go.jp/biz/areareports/2022/bf790ac1393fbb3c.html>〉

マートキーの生産に向けて工場建設を、フジテックはエクスプレス・リフツ（エレベーター工場を有する）の買収を、東レは水処理研究拠点の開設を、エア・ウォーターは液化ガス製造プラントの新設をそれぞれ発表した。自動車分野の研究開発では、ルネサスエレクトロニクスが財閥系タタ・エレクシーと協業し、電気自動車システム研究開発拠点を設立し、スズキも 100%出資の研究開発（R&D）センターを設立した。非製造業においては、富士通、メルカリ、楽天シンフォニー、NTT データ、PayPay が研究開発拠点を設置した¹⁸¹。

表 37 インドへの日本企業の進出状況

| 項目 | 主な内容 |
|----------|--|
| 日系企業の進出数 | 4,901 社（2022 年 10 月現在） ¹⁸² |
| 進出状況 | <ul style="list-style-type: none"> 進出企業の再投資が活発化（AGC、花王等の設備増強、NTT データの拠点開設、M&A でリニューアブル・ジャパンは太陽光発電所の取得、オリックスは、エラワンエナジーの完全子会社化を図った） |
| 在留邦人数 | 8,197 人（2023 年 10 月現在） ¹⁸³ |

（3）日本との関係

日本とインド両国は 1952 年に国交を樹立し、インド国内の強い親日感情もあり、友好関係を維持してきた。2000 年 8 月には「日印グローバル・パートナーシップ」構築に合意し、2005 年以降は、年次首脳会談を実施してきた。2014 年 9 月には、両国関係を特別戦略的グローバル・パートナーシップ（「日印特別戦略的グローバル・パートナーシップ」）へ格上げした。2023 年は日本が G7 議長国、インドが G20 議長国を担う年であり、両国の連携を確認するとともに、二国間関係を更に強化した。

2022 年 3 月の日印首脳会談では、今後 5 年間でインドへ 5 兆円（3 兆 2000 億ルピー）の投資を目指すことを発表した。投資は官民の両方で行われ、これには様々なプロジェクトに対するインドへの負債も含まれる。投資の大部分を呼び込むと見込まれる分野は、主に製造、気候変動（日印クリーンエネルギーパートナーシップ）、インフラ（高速鉄道建設事業）である¹⁸⁴。

（4）当該国の製造業の状況

インドの製造業市場は、グローバルプレイヤーとローカルプレイヤーが混在する断片的な市場である。市場に参入している主要企業には、Tata Motors Ltd、Mahindra Mahindra

¹⁸¹ 前掲：JETRO（2023）「インドの貿易と投資」

¹⁸² 外務省（2022）「海外進出日系企業拠点数調査」

¹⁸³ 外務省（2023）「海外在留邦人数調査統計」

¹⁸⁴ 外務省（2023）「日印首脳会談」, 2023 年 3 月 20 日.

〈https://www.mofa.go.jp/mofaj/s_sa/sw/in/page1_001542.html〉

Limited、Ashok Leyland、Hindustan Unilever Limited、Godrej group などがある。同市場の主要企業は、製造業界の顧客により良い製品とサービスを提供するため、主要な開発戦略として製品発表、提携、事業拡大、買収を採用している。例えば、2022年12月、タタ・モーターズの子会社であるタタ・パッセンジャー・エレクトリック・モビリティは、フォード・インドの車両製造工場を7億2,570万ルピー（8,901万米ドル）で買収した。この買収により、年産30万台、年間42万台まで拡張可能な最新鋭の製造能力が追加される¹⁸⁵。

インドのナレンドラ・モディ首相は、インドを製造業の拠点として世界地図に掲載し、インド経済を世界的に認知させるため、2014年より「メイク・イン・インドプログラム」を開始した。「メイク・イン・インドプログラム」は、製造業振興のスローガンであり、投資環境の整備を通じて、直接投資誘致を促進し、GDPに占める製造業の割合を15%から25%に引き上げる目標を掲げる。同プログラムでは、国内製造業保護と高付加価値の部品の国産化を推進する一方で、インド国内に進出する外資企業に向けては、多くの分野を開放、投資を積極的に誘致しており、特に、製造業の呼び込みには熱心である。法人税の引き下げや国内生産を促す生産連動型奨励金（PLI）スキーム等の優遇策を実施した。製造業振興と貿易赤字の解消を狙ったものである¹⁸⁶。

（5）地理的・地政学的重要性

①当該国と日本との友好関係

国交樹立以降、皇室－王室との交流のほか、外交面でも日本とインドは、民主主義や法の支配などの基本的価値や原則、また戦略的利益を共有するアジアの二大民主主義国であり、「日印特別戦略的グローバル・パートナーシップ」の下、経済、安全保障、人的交流など、幅広い協力を深化させてきた。

②経済安全保障やその他リスクに関する状況

インドは「自由で開かれたインド太平洋（FOIP）」を実現する上で重要なパートナーであり、日米豪印といった多国間での連携も着実に進展している。太平洋を臨む日本と、インド洋の中心に位置するインドが二国間及び多国間の連携を深めていくことは、インド太平洋の平和と繁栄に大いに貢献する。インド太平洋地域第2回日印外務・防衛閣僚会合（「2+2」）（9月8日、東京）の経済秩序の構築においてもインドは不可欠なプレイヤーであり、その意味でも地域的な包括的経済連携（RCEP）協定への将来的な復帰が期待される。

インドにおけるリスクとして、世界経済フォーラムの「エグゼクティブ・オピニオン調査（EOS）」は、デジタル・サービスの供給不足とデジタル格差（不平等）、戦略資源の地

¹⁸⁵ MordorIntelligence（2023）「インドの製造業市場規模」

〈<https://www.mordorintelligence.com/ja/industry-reports/india-manufacturing-sector-market>〉

¹⁸⁶ JETRO（2021）「メイク・イン・インドの成果に夜明け？」, 2021年4月22日.

〈<https://www.jetro.go.jp/biz/areareports/2021/3517a21c7c701096.html>〉

政学的な（覇権）争い、生活費危機、主要経済国の累積債務危機、自然災害と異常気象等を挙げた¹⁸⁷。

4.3.12. トルコ

(1) 当該国の概況

トルコでは、2023年5月に大統領選挙及びトルコ大国民議会議員選挙が行われ、エルドアン大統領が決選投票にて再選を果たし、2期目に入った（任期5年）。議会は、一院制であり、与党・共和同盟（公正発展党（AKP）、民族主義者行動党（MHP）等）は600議席のうち過半数超え323の議席を獲得し、議会の多数派となった。トルコは、2018年6月に議院内閣制から実権型大統領制に移行しており、内閣は大統領の任命によるものとなり、行政権の全てが大統領に属するようになった¹⁸⁸。

トルコ政府は、2023年9月に中期計画（2024～2026年）を発表し、低金利政策により投資、生産、雇用、輸出に基づく成長政策を通じた物価安定に重点を置き、インフレ問題を国の課題から取り除くことを掲げた。また、同計画では、経済成長と貿易、人材と雇用、物価と金融の安定、財政、災害管理、グリーンとデジタル化、ビジネスと投資環境に関するビジョンを提示した¹⁸⁹。

経済状況については、2022年は、新型コロナウイルス感染症の拡大下でのロックダウン（都市封鎖）の中にあっても生産活動を維持したこともあり、上半期に好調だった外需が雇用を支え、個人消費が成長をけん引した。また、ロシアによるウクライナ侵攻に対して、トルコは欧米の制裁に参加することなく、仲介姿勢をもって、エネルギーおよび商品物流の代替ルートを提示する等の独自路線もあり、実質GDP成長率は5.5%の高成長を実現した。一方で、低金利政策は、マイクロ経済の不均衡を拡大させており、消費者物価指数が前年比で64.3%まで高騰した¹⁹⁰。

(2) 日本企業の進出状況、ビジネス展開状況

日本企業の進出数は、2022年10月現在、275社である。日本企業がトルコに拠点を構えている理由として、市場の将来性、市場規模に続き、製造拠点としての優位性を挙げている¹⁹¹。

¹⁸⁷ World Economic Forum（2023）「第18回グローバルリスク報告書2023年版」。

¹⁸⁸ 外務省「国・地域」情報：トルコ共和国
〈<https://www.mofa.go.jp/mofaj/area/turkey/data.html>〉

¹⁸⁹ JETRO（2023）「2024～2026年の経済「中期計画」を発表、経済界は好意的に受け止め」、2023年9月19日。〈<https://www.jetro.go.jp/biznews/2023/09/4ed5cc4703013e3a.html>〉

¹⁹⁰ JETRO（2023）「トルコの貿易と投資」、2023年9月15日。
〈https://www.jetro.go.jp/world/middle_east/tr/gtir.html〉

¹⁹¹ JETRO（2023）「2023年度海外進出日系企業実態調査 中東編」、2023年12月21日。
〈https://www.jetro.go.jp/ext_images/_Reports/01/5911cd6f8d6e715c/20230029.pdf〉

トルコの国際収支統計では、2022年の日本のトルコ向け直接投資（株主資本インフロー）は、8,000万ドル（前年比64.0%減）と2009年以来最低の水準に冷え込んだ。直接投資案件には、トスヤルトーヨー（東洋鋼鋳とトスヤル・ホールディングの合弁会社）が、2022年8月にブリキ生産増強に向けた工場投資を発表し、2023年3月には同工場での太陽光発電施設の建設を発表した。また、自動車関連でも活発な動きがみられ、2022年4月にGSユアサがマニサ県の持分法適用関連会社IGYAの株式10%を追加取得し子会社化した。同年10月には、トヨタ・モーター・マニファクチャリング・ターキー（TMMT）が、ハイブリッド車（HV）とプラグインハイブリッド車（PHEV）バッテリーの新規生産投資を行うことを発表した。また、同月に、自動運転ソフトウェア開発のティアフォーが、自動運転スタートアップのレオドライブに出資した¹⁹²。2023年以降も、自動車分野の投資案件は続く（丸紅のテムサ・イシュ・マキナラルを完全子会社化（丸紅流通サービスに変更）、アナドルいすゞによるFZKエンジニアリングのカーカス生産事業の買収）。

表 38 トルコへの日本企業の進出状況

| 項目 | 主な内容 |
|----------|---|
| 日系企業の進出数 | 275社（2022年10月現在） ¹⁹³ |
| 進出状況 | <ul style="list-style-type: none"> 製造業拠点の優位性から製造業を中心に直接投資が行われている（トヨタ、GSユアサ、丸紅流通サービス、いすゞ） |
| 在留邦人数 | 1,752人（2023年10月現在） ¹⁹⁴ |

（3）日本との関係

日本とトルコとの関係は、1890年のエルトゥールル号事件（日本からの帰路に紀州・串本沖で座礁し、日本側が救護し、巡洋艦にて送還した事件）に遡る。友好関係は、現在にも続き、2011年3月に発生した東日本大震災では、トルコ政府は支援・救助チームを派遣し、他方、2011年10月のトルコ東部の地震には、救急援助物資を供与し、2023年2月のトルコ東南部で発生した地震に対しては、国際救急援助隊を派遣した。

2023年12月に開催した「日・トルコ首脳会談」では、イスラエルのパレスチナ自治区ガザへの侵攻について、深刻な人道上の懸念、戦闘休止等を表明した。また、2024年は、日本-トルコとの外交関係樹立100周年を迎え、経済、エネルギー、観光等の分野で二国間関係を更に発展させていくことで一致した¹⁹⁵。

（4）当該国の製造業の状況

トルコの製造業は、自動車産業では、生産台数135万台のうち、金額ベースで70%前後

¹⁹² 前掲：JETRO（2023）「スペインの貿易と投資」

¹⁹³ 外務省（2022）「海外進出日系企業拠点数調査」

¹⁹⁴ 外務省（2023）「海外在留邦人数調査統計」

¹⁹⁵ 外務省（2023）「日・トルコ首脳会談」, 2023年12月1日.

〈https://www.mofa.go.jp/mofaj/me_a/me1/tr/pageit_000001_00026.html〉

は、EU 向けであり、その他は中近東等の 100 カ国前後に輸出されている。軽商用車とバスは、EU 最大級の生産国である。トルコの純国産メーカーの TOGG による乗用電気自動車の生産販売を開始した¹⁹⁶。

中銀発表の国際収支統計によると、2022 年の対内直接投資（株主資本、フロー）は、67 億 9,100 万ドルであり、地域別では、EU からの投資額が全体の 67.4%を占める。業種別では、サービス部門のシェアが全体の 67.0%を占める。製造業では食品・飲料の投資額が前年の 3.5 倍、コンピュータ・電気・電子機器・精密機器が 81.5%増と好調だった。

（5）地理的・地政学的重要性

①当該国と日本との友好関係

前述の通り、日本とトルコとの関係は、1890 年に遡り、友好的な関係が構築されてきた。現在、日・トルコ経済連携協定交渉の早期妥結に向けて協議を続けている。トルコは、これまでに 20 以上の国・地域と FTA を締結しており、両国の経済界からも日・トルコ EPA の早期締結に対する高い期待感が示されていることから、2014 年 1 月の日・トルコ首脳会談において交渉開始に一致し、2022 年 12 月末までに 17 回の交渉会合が開催された¹⁹⁷。科学技術協力では、トルコ・日本科学技術大学（TJU）の開学等に取り組んでいる。

②経済安全保障やその他リスクに関する状況

欧州、中東、中央アジア・コーカサス地域、アフリカの結節点に位置する、地政学上重要な国であり、高い経済的潜在性を有し、周辺地域への輸出のための生産拠点としても注目されている。また、ロシアによるウクライナ侵略をめぐっては、両国との良好な関係をいかした積極的な仲介外交を展開しており、黒海を通じたウクライナからの穀物輸出再開の実現に大きく貢献した。外交面においては、引き続き、これまで関係の悪化が懸念されていた域内諸国との対話再開と関係の再構築が進められた¹⁹⁸。

トルコにおけるリスクとして、世界経済フォーラムの「エグゼクティブ・オピニオン調査（EOS）」は、急激な物価変動・長期化するインフレ、雇用や生活破綻（生活苦）の危機、国家間の紛争、著しい資源・物品供給の危機、主要経済国の累積債務危機、極端なコモディティショックや相場の乱高下等を挙げた¹⁹⁹。

¹⁹⁶ トルコ共和国大統領府投資局「競争力のある製造業」

〈<https://investturkey.or.jp/archives/63>〉

¹⁹⁷ 外務省「日・トルコ経済連携協定」

〈https://www.mofa.go.jp/mofaj/gaiko/fta/j_turkey/index.html〉

¹⁹⁸ 前掲：外務省「国・地域」情報：トルコ共和国

¹⁹⁹ World Economic Forum（2023）「第 18 回グローバルリスク報告書 2023 年版」。

4.4. 調査対象地域（ASEAN、中南米、中東等）の総括

4.4.1. 海外における知的財産活用支援の枠組み

欧州知的財産ヘルプデスク（The European IP Helpdesk）は、欧州委員会の域内市場・産業・起業・中小企業総局（DG Grow）が政策指導を行い、欧州イノベーション会議・中小企業執行機関（EISMEA）が運営するサービスである。

EISMEA は、EU レベルで技術やその他の知的財産権、知的財産資産を管理、普及、価値化するための国境を越えた中小企業や研究活動を支援している。広範な情報資料、知的財産を直接支援するヘルプライン・サービス、オンサイトおよびオンライン・トレーニングなどを提供する欧州知的財産ヘルプデスクの主な目標は、知的財産に関する認識から戦略的利用、活用の成功に至るまで、知的財産実務の全段階に沿った知的財産能力の構築を支援することである。この知的財産能力の強化は、EU の中小企業、EU が資金提供するプロジェクトの参加者・候補者、EU のイノベーション関係者に焦点を当て、EU のイノベーション・エコシステムへの知的財産の導入を促進することを目的としている。特定の知的財産問題について個人的なサポートが必要な場合、欧州における知的財産とイノベーションの世界の最新動向について情報を得たい場合、知的財産に関する研修に興味がある場合など、欧州知的財産ヘルプデスクは最適なパートナーといえる²⁰⁰。

(1) IP リソースライブラリ

IP リソースライブラリでは、ガイド、公報、ファクトシート、ケーススタディなど、多面的な IP および IP マネジメントの世界の様々なトピックに取り組んだ、実践的でユーザーフレンドリーで読みやすい様々な出版物を閲覧できる。

IP resources library

In our IP resource library you will find a variety of hands-on, user friendly and easy-to-read publications such as guides, Bulletins, fact sheets or case studies tackling different topics from the multifaceted world of IP and IP management.



図 24 IP resources library の構成

出所：欧州知的財産ヘルプデスク（The European IP Helpdesk）

https://intellectual-property-helpdesk.ec.europa.eu/regional-helpdesks/european-ip-helpdesk_en

²⁰⁰ Our services : <https://www.iprhelpdesk.eu/home>

(2) 主要地域別のヘルプデスク

欧州知的財産ヘルプデスクでは、次図のように中国、欧州、インド、ラテンアメリカ、東南アジア、アフリカといった主要地域別にカテゴリ化して情報提供している。

Intellectual Property Helpdesk



図 25 Intellectual Property Helpdesk の主要地域別サービス

出所：欧州知的財産ヘルプデスク（The European IP Helpdesk）

https://intellectual-property-helpdesk.ec.europa.eu/index_en

以下、本調査の対象に含まれているインド、ラテンアメリカ、東南アジアについて概略を紹介する。

①インド

India IP SME Helpdesk は、欧州連合(EU)および単一市場プログラム(SMP)関連国²⁰¹の中小企業が、インドにおいて、またはインドに関連する知的財産権(IP)を保護・行使できるよう、無料の情報・サービスを提供している。これらのサービスには、知的財産権や関連問題に関する専門用語を使わない第一線の秘密厳守のアドバイスや、研修、資料、オンライン・リソースが含まれる。

Helpline

Our Intellectual Property experts provide professional, confidential and tailor-made answers to your IP-related questions within a maximum of 3 working days.

What is the Helpline

Free of charge, tailor-made helpline

Book a meeting

Request a meeting with our IP experts

Case Studies

Case studies from users of the India IP SME Helpdesk

FAQ

Most commonly requested information about IP and India

図 26 India IP SME Helpdesk における Helpline

²⁰¹ 単一市場プログラムの対象国は、現在すべての EU 加盟国に加え、ノルウェー、アイスランド、リヒテンシュタイン、トルコ、ウクライナ、北マケドニア、コソボ、ボスニア・ヘルツェゴビナである。

出所：欧州知的財産ヘルプデスク（The European IP Helpdesk）

https://intellectual-property-helpdesk.ec.europa.eu/regional-helpdesks/india-ip-sme-helpdesk_en

② ラテンアメリカ

Latin America IP SME Helpdesk は、ラテンアメリカ諸国（アルゼンチン、ボリビア、ブラジル、チリ、コロンビア、コスタリカ、キューバ、エクアドル、エルサルバドル、グアテマラ、ホンジュラス、メキシコ、ニカラグア、パナマ、パラグアイ、ペルー、ウルグアイ、ベネズエラ）における知的財産権の保護と権利行使のために、欧州連合（EU）および単一市場プログラム（SMP）加盟国の中小企業を支援し、無料の情報とサービスを提供している。これらのサービスには、知的財産権および関連問題に関する、専門用語を使わない第一線の秘密厳守のアドバイスや、研修、資料、オンライン・リソースが含まれる。

Helpline

Our Intellectual Property experts provide professional, confidential and tailor-made answers to your IP-related questions within a maximum of 3 working days.

| | |
|---|---|
| Register and ask Free of charge, tailor-made helpline | Case studies Case studies from users of the Latin America IP SME Helpdesk |
| Testimonials What do they say about us? | FAQ Most commonly requested information about IP and Latin America |

図 27 Latin America IP SME Helpdesk における Helpline

出所：欧州知的財産ヘルプデスク（The European IP Helpdesk）

https://intellectual-property-helpdesk.ec.europa.eu/regional-helpdesks/latin-america-ip-sme-helpdesk_en

IP Library

Open-access IP materials for EU SMEs doing business in Latin America.

| | | |
|---|---|---|
| IP yearly reviews Yearly summaries of the main IP developments in Latin America | Country materials Country-focused IP materials | IP basics Learning resources on IPRs in Latin America |
| Green sector IP materials on the green tech sector | Digital Transformation IP materials on the ICT sector | Other sectors IP materials on other sectors (e.g. Agrifood, E-Commerce, Health) |
| Blog Monthly articles on IP in Latin America | Useful documents Templates (e.g. NDAs) | Advanced search All IP Helpdesk Publications in one place |

図 28 Latin America IP SME Helpdesk における IP resources library

出所：欧州知的財産ヘルプデスク（The European IP Helpdesk）

https://intellectual-property-helpdesk.ec.europa.eu/regional-helpdesks/latin-america-ip-sme-helpdesk_en

※Training の構成はインドと同様のため割愛。

③ 東南アジア

South-East Asia IP SME Helpdesk は、欧州連合（EU）および単一市場プログラム（SMP）に加盟する他の国々の中小企業が、東南アジア諸国において、または東南アジア諸国に関連する知的財産権（IP）を保護・行使できるよう、無料の情報およびサービスの提供を通じて支援する。これらのサービスは、知的財産権および関連する問題についての専門用語を使わない、第一線の秘密厳守のアドバイス、さらに研修、資料、オンライン・リソースの形で提供される。

Helpline

The South-East Asia IP SME Helpdesk is there to help you. Send us your enquiry and you will receive a response within 3 working days.

[Helpline - submit enquiry](#)

[IP Institutions Directory](#)

図 29 South-East Asia IP SME Helpdesk における Helpline

出所：欧州知的財産ヘルプデスク（The European IP Helpdesk）

https://intellectual-property-helpdesk.ec.europa.eu/regional-helpdesks/south-east-asia-ip-sme-helpdesk_en

IP Resources library

Use this database to find the resource you are looking for.

[Country Factsheets](#)

[IP Guides](#)

[Case studies](#)

[Infographics](#)

[Podcasts](#)

[Videos](#)

[Glossary](#)

図 30 South-East Asia IP SME Helpdesk における IP resources library

出所：欧州知的財産ヘルプデスク（The European IP Helpdesk）

https://intellectual-property-helpdesk.ec.europa.eu/regional-helpdesks/south-east-asia-ip-sme-helpdesk_en

IP resources library

This library provides open-access IP materials especially designed for EU SMEs doing business in India.

| | | |
|--|---|---|
| India key content CFS, EU comparatives, Factograph | IP basics Learning resources on IPRs in India | Blog Monthly articles on IP in India |
| Green Sector IP materials on the green tech sector | Digital Transformation IP materials on the ICT sector | Other Sectors IP materials on other sectors (e.g. Agrifood, E-Commerce, Health) |
| Useful documents Templates (e.g. NDAs) | Advanced search All IP Helpdesk Publications in one place | |

Training

The Helpdesk offers on-site and online [training activities](#) that cover relevant IP topics for EU SMEs and entrepreneurs looking to expand their businesses to India.

| | | |
|--|---|--|
| Upcoming trainings Join one of our future webinars or onsite trainings | E-Learning Try one of our recorded training modules | Organise a training with us! Tailor-made to the needs of the business partners |
|--|---|--|

図 31 India IP SME Helpdesk における IP resources library および Training

出所：欧州知的財産ヘルプデスク（The European IP Helpdesk）

https://intellectual-property-helpdesk.ec.europa.eu/regional-helpdesks/india-ip-sme-helpdesk_en

4.4.2. Country Factsheets における特許に係る言語情報

欧州知的財産ヘルプデスクで示された Country Factsheets の中には、欧州の中小企業が当該国で特許出願する際の要件等の基礎情報が記載される。東南アジアの Country Factsheet では、言語情報に関する記載、特許出願に係るヒントや注意事項の記載が含まれる。

以下、Country Factsheet における言語情報を取りまとめた。これらの国々での出願においては、現地語での出願が求められる。

表 39 Country Factsheet における言語に関する情報

| | 特許出願に係る言語情報 | 特許出願に係るヒント |
|------|------------------------------------|--|
| インド | (記載なし) | |
| メキシコ | (記載なし) | |
| ブラジル | 申請書およびその他の必要書類はポルトガル語で提出 | ブラジル法では、特許および実用新案について「グレースピリオド（猶予期間）」を設けている。 |
| タイ | 申請書およびその他の必要書類（関連する場合は）、タイ語で提出するか、 | 現在、タイで出願される特許出願のほとんどは、PCT ルートで出願されている |

| | 特許出願に係る言語情報 | 特許出願に係るヒント |
|--------|--|---|
| | タイ語の翻訳文を添付しなければならない（特許明細書、委任状、譲渡証書など）。英文明細書のタイ語翻訳、公証済み委任状、譲渡証書（出願人が発明者でない場合）または出願人の権利に関する声明書（出願人が発明者である場合）は、タイへの出願日から 90 日以内に提出可能。この期限は延長できない。 | る。タイでの PCT ルートには多くの利点があるが、主な利点は、タイでの実体審査請求に要する時間の間に、対応する外国特許出願が外国で付与されている可能性がある |
| ベトナム | 委任状、証拠書類（特に登録権及び優先権に関するもの）、その他出願をサポートする書類を除き、出願書類はすべてベトナム語で作成しなければならない。IP ベトナムの要請によりベトナム語に翻訳しなければならない。 | 特許権侵害事件に対する行政執行機関の対応は、これまであまり効果的でなかったため、特許権侵害に対しては、行政措置ではなく民事措置をとることが望ましい。 |
| インドネシア | 特許出願はインドネシア語（Bahasa Indonesia）で行わなければならない。ただし、特許請求の範囲および明細書が英語で出願された場合は、出願日から 1 ヶ月以内にインドネシア語に翻訳したものを提出することが認められる。 | 特許明細書は現地の言語に翻訳する必要があるが、翻訳ミスはよくあることで、翻訳が直訳的すぎることも多い。特許が付与された後にこのような誤りを訂正する手続がない。事態の悪化を避けるため、独立請求項は特許明細書の中で最も重要な部分であるため、出願人は独立請求項の翻訳に特別な注意を払うことを推奨する。 |
| マレーシア | 提出書類はすべて英語またはマレーシアの国語（Bahasa Malaysia）で書かれたものでなければならない。 | 出願人は、マレーシアにおける特許登録手続を短縮するために、PPH または ASPEC の早期審査プログラムの利用を検討することができる。 |

出所：欧州知的財産ヘルプデスク（The European IP Helpdesk）の Country Factsheet より作成。

4.4.3. ASEAN、中南米、中東における民間のニーズ

本調査では、ASEAN、中南米、中東地域で事業活動を展開する民間企業を対象に書面調査を実施した。調査対象の事業者は、当該地域において特許出願を行っている社を候補に打診を行い、3社を対象に実施した。

表 40 調査対象企業

| 本報告書略記 | 業種 |
|--------|------|
| 企業 A | 電気機器 |
| 企業 B | 電気機器 |
| 企業 C | 電気機器 |

民間ニーズの調査項目は、(1) ASEAN/南米/中東地域での特許・技術調査経験及び特許出願・権利化経験、(2) ASEAN/南米/中東地域の知財庁への日本国特許庁（JPO）の期待

について、現状の取組みと要望に関する情報収集を行った。

(1) ASEAN/南米/中東地域での特許・技術調査経験及び特許出願・権利化経験

ASEAN/南米/中東地域でのビジネス展開／特許出願・権利化にあたって、各種調査の実施状況については、FTO（侵害予防調査）を実施している社は、1社であった。また、先行技術調査を実施している社も1社であった。当該地域について、両方の調査を実施している社は見られなかった。

表 41 特許出願・権利化にあたっての各種調査の実施状況

| | FTO（侵害予防調査）の実施 | 先行技術調査 |
|-----|----------------|---------|
| 企業A | 実施している | 実施していない |
| 企業B | 実施していない | 実施していない |
| 企業C | わからない | 実施している |

なお、実施されている調査において、現地語の翻訳文の精度、データベースの収録範囲等、課題と感じられる点については、下記を挙げた。

- FTO 調査については日米欧中等の主要国の調査を原則とし、事業実施国におけるファミリー出願の有無を確認している。このため、現地国・地域の特許権そのものを直接確認することは稀である。重要性が高い場合には調査を実施することはある。商用データベースにおける公報の収録範囲が十分であるか、英語未対応の現地庁サイトの使いにくさ、機械翻訳の精度が課題となる。（企業A）
- ASEAN 地域については、IP ランドスケープ調査を行い、現地のプレイヤーや顧客候補企業の調査を行うことがあるが、広い年代にわたって幅広く網羅しているデータベースがないため、データが偏ってしまう。中東はイスラエルのみ調査対象とすることがあるが、通常のデータベースで大きな問題は感じていない。南米は調査していない。（企業C）

ASEAN/南米/中東地域での特許出願・権利化にあたって、各知財庁での審査・運用面で課題と感じられる点については、下記を挙げた。

- 出願時に“原紙”の提出が要求される書面（サインフォーム等）が存在するため、電子化されることを望む。（企業A）
- 特に課題と感じている点はない。ブラジルの審査期間は改善された印象を持っている。（企業B）
- 「特許出願・権利化」という目的での調査は行っていないが、時々、権利の維持状況を調査する必要が生じる。しかし、著名なデータベースであってもステータスが収録されていないことが多く、現地特許庁のシステムにアクセスする必要に迫られること

があるが、言語の壁も含めて利用のハードルが非常に高く、目的とする知財にたどり着けないことも多い。(企業 C)

ASEAN/南米/中東地域での特許出願・権利化にあたって、日本語から出願言語への翻訳上の課題と感じられる点については、下記を挙げた。

- 作成された翻訳が真に正しいのか検証しづらいことが課題である。翻訳精度の関係で日本語→英語→現地語の順に翻訳することがある。英語圏への出願も行う場合には良いが、そうでない場合には翻訳費用の増大に繋がるため、予算面での課題が生じる。(企業 A)
- 「特許出願・権利化」に関与していないので、翻訳上の問題は不明。(企業 C)

(2) ASEAN/南米/中東地域の知財庁への日本国特許庁 (JPO) の期待

日本国特許庁に関する期待は、当該地域における知財制度の実効性の向上や、特許審査の質の維持・向上、知財庁のデータ集約し EPO に代行送付できるような機能を期待した。

- 各国での知財制度の実効性の向上（適切な権利取得・適切な権利行使ができる環境の整備による予見性の向上）に向けて取り組んで頂けることを期待したい。(企業 A)
- 今後も当該地域の特許庁とコミュニケーションを図り、特許審査の質の維持、向上を図ってほしい。(企業 B)
- 現地特許庁のデータ集約し、データのクレンジングを行い、EPO に代行送付（現地特許庁が EPO にデータを送るにあたって何が問題なのかを解決してスムーズに送られるようにサポートする（極端な例が代行）(企業 C)

(3) 当該地域の翻訳ニーズ（その他）

ヒアリング調査の実施にあたり、先行調査として、当該地域の特許情報の利用動向と翻訳ニーズについて、別の電気機械企業を対象に書面調査を実施した。

先行調査企業からは、ASEAN での特許情報の利用実績はあるが、非定期的な利用形態であることが示された。また、中南米、中東地域の特許情報は今後も利用する機会がない可能性が示された。当該地域の翻訳ニーズについては、市販の特許調査ツールで英文翻訳された情報を利用していることと、現地語の公報をテキストで入手することができないので翻訳機会がないとした。特許翻訳に係る課題として、翻訳の精度などいくつかの課題や期待はあるとした。

5. 調査対象言語に関するコーパス作成の技術的情報

5.1. 調査の概要

本章では調査対象とした 10 言語について、特に特許情報に係るコーパス作成の技術的情報として、パテントファミリーを利用したコーパス作成の具体的な候補手順とそれらの課題について報告する。特に、以下の点に分けて、言語ごとにまとめる。

5.2. パテントファミリーデータの取得元およびデータの状態

5.3. 多言語 OCR ツール

5.4. 文分割ツール

5.5. 形態素解析（単語分割）ツール

5.6. アライメントツール

5.7. 対訳辞書作成ツール

5.8. パテントファミリーデータからの対訳コーパス作成手順

5.9. コーパス試作、精度評価、作成可能な対訳コーパス量

エラー! 参照元が見つかりません。 . 大規模対訳コーパス作成の推定コスト

5.10. その他の関連技術

5.2. パテントファミリーデータの取得元およびデータの状態

対象となる 10 言語の公報データの入手に関し、人手によって公報データの取得を試みた際の手順と結果について以下に記載する。

5.2.1. スペイン語

スペイン語を公用語とする国・地域は複数あるが、ここではスペインとメキシコについて調査した。

スペイン王国

スペイン特許商標庁の検索ウェブサイトまたは Espacenet スペイン版等からスペイン公報を検索して公報データを取得することができる。

<手順>

- ・ スペイン特許商標庁（Invenes: OEPM）の英語版検索ウェブサイト²⁰²にアクセスし、「Advanced search」より所定の条件で公報を検索することができる。例えば「Search in title」で発明の名称に含まれるキーワード（例 Engine）を入力して「SEARCH」をクリックすると検索結果一覧が表示される。
- ・ 一覧に PDF のアイコンが表示されている公報については明細書、請求項の透明テキスト付き PDF およびテキストが取得可能（大半は PDF およびテキストがあるので

²⁰² <https://consultas2.oepm.es/InvenesWeb/faces/busquedaInternet.jsp>

問題ないと思われる。)。 「REFERENCE」 行の番号部分をクリックし、詳細画面を表示させて「Publication number」の公開番号をクリックすることにより明細書、請求項の透明テキスト付き PDF を取得できる。テキストについては詳細画面の「DESCRIPTION」をクリックすると明細書、「CLAIMS」をクリックすると請求項のテキストが取得できる。

パテントファミリーに日本公報が含まれるものに限ってスペイン公報のデータを取得する方法としては以下が挙げられる。

- Patbase 等の商用の特許検索・特許調査ツールを使用し、パテントファミリーに日本公報を含むスペイン公報の番号リストを出力し、特定されたスペイン公報の番号を使用して OEPM の検索ウェブサイトから公報データを取得する。

<特記事項>

- OEPM の検索ウェブサイトから優先権主張番号が日本の出願番号であるものに限って公報を検索することはできなかった。
- 要約は英語のみが記載されている公報が多く、スペイン語の要約がないものが多いと思われる。
- スペイン公報の大半が透明テキスト付き PDF だが、Word 等のファイルにコピーしたときに頁番号、行番号、および不要な改行等が多数挿入されるため、PDF からテキストを抽出する場合は、コーパス作成の前に不要な文字列や改行記号を削除する作業が必要となる。
- Espacenet スペイン版²⁰³に直接アクセスして Smart search や Advanced search 等から検索し、公報全文 PDF やテキスト（要約、明細書、請求項）を取得することも可能。同じく、LATIPAT²⁰⁴での検索も可能だが、OEPM の検索ウェブサイトの方がスペイン公報に絞った検索がしやすいと思われる。LATIPAT からの検索については、次のメキシコ合衆国で紹介する。

メキシコ合衆国

メキシコ産業財産庁の検索ウェブサイトまたは LATIPAT 等からメキシコ公報を検索して公報データを取得することができる。

<手順>

- メキシコ産業財産庁（IMPI）の検索ウェブサイト²⁰⁵にアクセスし、公報検索ができると思われるが、調査時はプルダウンメニューや検索ボタンがすべてグレーアウトされており操作できなかった。

²⁰³ https://es.espacenet.com/?locale=es_ES

²⁰⁴ Espacenet の技術による、ラテンアメリカ（スペイン語、ポルトガル語圏）およびスペインの特許文献を検索するためのオンラインサービス

²⁰⁵ <https://siga.impi.gob.mx/>

- ・ 次に LATIPAT²⁰⁶での手順を紹介する。英語版に変更したい場合は右上の「English」をクリックする。
- ・ 「Smart search」から、例えば「MX 2018 motor（出願国：メキシコ、公報発行年：2018年、キーワード：エンジン）」で検索をかけると、これらの条件をすべて満たすスペイン語公報が検索結果一覧に表示されるので、閲覧したい公報をクリックし、その公報詳細画面を表示させる。
- ・ 「Original document」をクリックすると画面中央に「Download」があるのでダウンロードするとメキシコ公報全文 PDF（要約、明細書、請求項）が取得できる。PDFからはテキストが抽出できないが、Espacenet 英語版²⁰⁷と同様に、詳細画面の「Bibliographic data」、「Description」、「Claims」をクリックするとそれぞれテキストが取得できる。

<特記事項>

- ・ LATIPAT は Smart search の他に、「Advanced search」等でも検索可能。20年以上前の公報も公報全文 PDF やテキスト（要約、明細書、請求項）が取得できるものが多い。

5.2.2. ポルトガル語

ポルトガル語を公用語とする国・地域は複数あるが、ここではブラジルとポルトガルについて調査した。

ブラジル連邦共和国

ブラジル工業所有権機関の検索ウェブサイトまたは LATIPAT 等からブラジル公報を検索して公報データを取得することができる。

<手順>

- ・ ブラジル工業所有権機関（INPI）の検索ウェブサイト²⁰⁸にアクセスし、所定の条件で公報を検索することができる。
- ・ 最初の画面で ID とパスワードを求められるが、全て空欄のままの状態です。「Continuar（英語で Continue の意）」をクリックして検索画面に進むことができる。
- ・ 次の画面で「Patente（英語で Patent の意）」をクリックして特許の検索ページに入る。
- ・ 例えば「Contenha（英語で Contains の意）」の入力枠を使用してキーワード検索を行う場合、「todas as palavras（英語で All the words の意）」と「Titulo（英

²⁰⁶ <https://lp.espacenet.com/>

²⁰⁷ https://worldwide.espacenet.com/?locale=en_EP

²⁰⁸ <https://busca.inpi.gov.br/pePI/>

語で Title の意)」の間にある入力枠に発明の名称のキーワード（例 motor、ポルトガル語と英語共に同じ語）を入れ、「pesquisar（英語で Search の意）」をクリックすると検索結果一覧が表示される。

- ・ 該当公報が表示された場合は詳細画面に入り、最下部にある公報表紙と思われるアイコンをクリックすることによって公報全文 PDF（要約、明細書、請求項）を取得できる。

パテントファミリーに日本公報が含まれるものに限ってブラジル公報のデータを取得する方法としては以下が挙げられる。

- ・ Patbase 等の商用の特許検索・特許調査ツールを使用し、パテントファミリーに日本公報を含むブラジル公報の番号リストを出力し、特定されたブラジル公報の番号を使用して INPI の検索ウェブサイトから公報データを取得する。

<特記事項>

- ・ INPI の検索ウェブサイトから優先権主張番号が日本の出願番号であるものに限って公報を検索することはできなかった。
- ・ ブラジル公報の大半が透明テキスト付き PDF であるためテキストの取得が可能ではあるが、Word 等のファイルにコピーしたときに頁番号、行番号、および不要な改行等が多数挿入されるため、コーパス作成の前に不要な文字列や改行記号を削除する作業が必要となる。
- ・ ブラジルも LATIPAT に参加しているため、メキシコ同様、ブラジル公報の検索ができ、公報全文 PDF、テキスト（要約、明細書、請求項）も取得できる。

ポルトガル共和国

ポルトガル工業所有権機関の検索ウェブサイト等からポルトガル公報を検索して公報データを取得することができる。

<手順>

- ・ ポルトガル工業所有権機関（INPI）の英語版検索ウェブサイト²⁰⁹にアクセスし、所定の条件で公報を検索することができる。
- ・ Registration Number、Applicant/Owner、Title、Abstract の 4 つの検索条件から選べるが、例えば「By Title」にチェックを入れ、下の「Next」をクリックするとキーワード検索ができるので、発明の名称のキーワード（例 motor、ポルトガル語と英語共に同じ語）を入れ、「Search」をクリックすると検索結果一覧が表示される。（なお、キーワード検索はポルトガル語のみ対応している。）
- ・ 優先日が古い順に表示されるので、新しい公報を閲覧したい場合はページ送りする。また、特許以外に実用新案等も一覧に出てくるので左列に「National Patent」

²⁰⁹ <https://servicosonline.inpi.pt/pesquisas/main/patentes.jsp?lang=EN>

と書かれているものか確認する。

- ・ 閲覧したい公報の Registration Number をクリックすると詳細画面が表示されるので、画面最下部の「Search patent in Esp@cenet」をクリックすると Espacenet ポルトガル版へ画面が移り、その公報のみが表示されるので、クリックし詳細画面を表示させる。Espacenet 英語版等と同じレイアウトになっているので、同様の方法で公報全文 PDF やテキスト（要約、明細書、請求項）が取得できる。

<特記事項>

- ・ Espacenet ポルトガル版²¹⁰に直接アクセスして Smart search や Advanced search 等から検索し、公報全文 PDF やテキスト（要約、明細書、請求項）を取得することも可能。ただし、LATIPAT には参加していない。

5.2.3. イタリア語

イタリア共和国

イタリア特許商標庁の検索ウェブサイトや Espacenet イタリア版等からイタリア公報を検索して公報データを取得することができる。

<手順>

- ・ イタリア特許商標庁 (National Patent Database: UIBM) の検索ウェブサイト²¹¹にアクセスし、所定の条件で公報を検索することができる。
- ・ 例えば「Con tutte le parole (英語で With all the words の意)」の入力枠を使用してキーワード検索を行う場合、この入力枠に発明の名称のキーワード (例 motore (英語で motor の意)) をイタリア語で入力し、「Avvia ricerca (英語で Start Search の意)」をクリックすると検索結果一覧が表示される。
- ・ 表示された公報の中から閲覧したい公報を選択し、「Numero deposito (英語で Deposit number の意)」欄の「10～の数字から始まる番号 (例 102013902219357)」をコピーする。
- ・ 続けてその下にあるメニューの「Dettagli (英語で Details の意)」をクリックし、さらに下部にある「Collegamento a DB Nazionale: (英語で Link to National DB の意)」欄の「Banca_dati (英語で database の意)」をクリックすると別のデータベースに移動する。
- ・ 表示された入力枠にコピーしておいた番号を入れて「Cerca (英語で Search の意)」をクリックするとその公報の詳細が表示される。詳細情報の下部にある「Documentazione della domanda depositata (英語で Documentation of the application field の意)」をクリックすると明細書、請求項の透明テキスト付き PDF が取得できる。

²¹⁰ https://pt.espacenet.com/?locale=pt_PT

²¹¹ <http://brevettidb.uibm.gov.it/>

パテントファミリーに日本公報が含まれるものに限定してイタリア公報のデータを取得する方法としては以下が挙げられる。

- ・ Patbase 等の商用の特許検索・特許調査ツールを使用し、パテントファミリーに日本公報を含むイタリア公報の番号リストを出力し、特定されたイタリア公報の番号を使用して UIBM の検索ウェブサイトから公報データを取得する。

<特記事項>

- ・ UIBM の検索ウェブサイトから優先権主張番号が日本の出願番号であるものに限定して公報を検索することはできなかった。
- ・ イタリア語の要約がない公報が多くあると思われる。
- ・ 公報 PDF の中には請求項のみイタリア語と英語の両方が併記されているものがあり、その場合、コーパス作成の際には英語部分を削除する必要がある。
- ・ 「Collegamento a DB Nazionale (英語で Link to National DB の意)」欄の下に「Collegamento a Espacenet (英語で Link to Espacenet の意)」欄があり、「Visualizza scheda (英語で View Tab の意)」をクリックすることで Espacenet のその公報のページに移動することができる。(Espacenet にその公報が表示される場合は、テキスト(明細書、請求項)が取得できるものもある)。
- ・ UIBM の検索ウェブサイトから検索できる公報の範囲は 2008 年 7 月から 2015 年 6 月までに出版された公報のみであり、検索結果は最新のものから上限 1000 件までが表示される仕様になっている。
- ・ Espacenet イタリア版²¹²に直接アクセスして Smart search や Advanced search 等から検索し、明細書、請求項の PDF やテキストを取得することも可能。

5.2.4. インドネシア語

インドネシア共和国

インドネシア知的財産総局の検索ウェブサイトや ASEAN IP REGISGER 等からインドネシア公報を検索して公報データを取得することができる。

<手順>

- ・ インドネシア知的財産総局 (Director General of Intellectual Property: DGIP) の検索ウェブサイト²¹³にアクセスし、所定のメニューからデータベースのウェブページ²¹⁴に移動する。
- ・ 表示された画面において特許、商標、意匠、著作権等のリストの中から「Patent (「特許」の意)」を選択し、検索対象となる出願番号を入力して検索する(出願

²¹² <https://it.espacenet.com/>

²¹³ <http://www.dgip.go.id/>

²¹⁴ <https://pdki-indonesia.dgip.go.id/>

番号の例：P00202200103)。

- ・ 検索結果から該当公報が表示されたらクリックして公報ページを開き、画面右「Publikasi (英語で Publication の意)」欄のダウンロードボタンをクリックし、組織名、氏名、メールアドレス、電話番号を入力すると透明テキスト付きの公報全文 PDF (要約、明細書、請求項) が取得できる。

パテントファミリーに日本公報が含まれるものに限定してインドネシア公報のデータを取得する方法としては以下が挙げられる。

- ・ Patbase 等の商用の特許検索・特許調査ツールを使用し、パテントファミリーに日本公報を含むインドネシア公報の番号リストを出力し、特定されたインドネシア公報の番号を使用して DGIP の検索ウェブサイトから公報データを取得する。
- ・ DGIP の検索ウェブサイトの「Advance Filter」のメニューから「Nomor Prioritas (英語で Priority Number の意)」欄に「JP」と入力し、優先権主張番号が日本の出願番号であるものに限定して公報検索を行って公報データを取得する。
- ・ ASEAN IP REGISGER からの検索方法は次の通りである。ASEAN IP REGISGER 日本語版²¹⁵より、インドネシアの「特許」列の青い数字部分をクリックする。上部の「詳細」をクリックすると、詳細検索画面になるので、上部右にある「フィールドの追加/削除」をクリックして、「国」のオプションメニューから「優先権の詳細」にチェックを入れると、画面中央に入力欄が表示される。「JP」と入力し、「検索」をクリックすると検索結果一覧が表示されるので、閲覧したい公報をクリックすると、その公報の詳細画面から要約のテキストが取得できる。また、公報によっては明細書、請求項のテキストを取得できるものもある。他にも色々なフィールドを追加することができるので、条件を絞って検索することができる。

<特記事項>

- ・ 工業所有権情報・研修館 (INPIT) のウェブサイトである「新興国等知財情報データベース」の「インドネシアにおける特許・簡易特許 (実用新案) 公報のアクセス方法」²¹⁶に記載された手順を参照するとよい。
- ・ 公報全文 PDF について、ANNOUNCEMENT DATE の 2018 年と 2019 年辺りでは全文が取得できるものとできないものが混在しているが、調査した結果において、それ以降は全文が取得できるものが多いと思われる。また、2000 年頃の公報 (出願番号の例：P00200000104) で確認したところ、公報全文 PDF が取得できない公報でも、要約や一部の書誌情報はテキストで掲載されているものが多い。
- ・ インドネシア公報の大半が透明テキスト付き PDF であるためテキストの取得が可能ではあるが、Word 等のファイルにコピーしたときに頁番号、行番号、および不要な改行等が多数挿入されるため、コーパス作成の前に不要な文字列や改行記号

²¹⁵ <https://asean-ipregister.wipo.net/wopublish-search/public/home?13&lang=jp>

²¹⁶ <https://www.globalipdb.inpit.go.jp/etc/19688/>

を削除する作業が必要となる。

5.2.5. マレー語

マレーシア

マレーシア知的財産公社の検索ウェブサイトや ASEAN IP REGISGER 等からマレーシア公報の検索は可能であるものの、公報データの取得はできないものが多い。

<手順>

- ・マレーシア知的財産公社の検索ウェブサイト (IP ONLINE: MyIPO)²¹⁷にアクセスし、所定の条件で公報を検索することができる。
- ・「Search Now」をクリックし、「Patents」欄の「Search」をクリックすると検索画面が表示される。
- ・「Advanced Search」をクリックすると詳細検索ができる。例えば、「Abstract」のキーワード検索でマレー語である「enjin (英語で engine の意)」で検索して一番下の「Divisional Application」欄にある「Search」をクリックするとマレー語公報の検索結果一覧が表示される。マレー語と英語が同じ単語や英語でキーワード検索すると英語の公報が検索結果一覧に多く出てくる。また、「Priority Country (日本語で優先国)」欄のプルダウンメニューから「Japan」を選んで検索すると英語の公報ばかりが検索結果一覧に出てくる。
- ・ASEAN IP REGISGER からの検索方法は次の通りである。ASEAN IP REGISGER 日本語版²¹⁸より、マレーシアの「特許」列の青い数字部分をクリックする。上部の「詳細」をクリックすると、詳細検索画面になるので、上部右にある「フィールドの追加/削除」をクリックして、「国」のオプションメニューから「優先権の詳細」にチェックを入れると、画面中央に入力欄が表示される。「JP」と入力し、「検索」をクリックすると検索結果一覧が表示されるので、閲覧したい公報をクリックすると、その公報の詳細画面から要約のテキストが取得できる。また、他にも色々なフィールドを追加することができるので、条件を絞って検索することができる。調査時においては、明細書、請求項のテキストは見つけれず、MyIPO と同様、検索結果一覧を見る限り、すべて英語の公報になっておりマレー語の公報は見つけれなかった。

<特記事項>

- ・閲覧したい公報の詳細画面に入ることができても公報全文のデータを取得できないものが多いと思われるが、マレー語の要約は取得可能。

5.2.6. ベトナム語

ベトナム社会主義共和国

²¹⁷ <https://iponline2u.myipo.gov.my/PT.html>

²¹⁸ <https://asean-ipregister.wipo.net/wopublish-search/public/home?13&lang=jp>

ベトナム国家知的財産庁の検索ウェブサイト等からベトナム公報を検索して公報データを取得することができる。

<手順>

- ・ ベトナム国家知的財産庁 (IP VIET NAM) の検索ウェブサイト²¹⁹にアクセスし、所定のメニューからデータベースのウェブページ (日本語版)²²⁰に移動する。
- ・ 「出願番号」欄に検索対象となる出願番号を入力して「検索」をクリックする (出願番号の例: 1-2021-07157)。
- ・ 該当公報が表示されたらクリックして公報ページを開き、「ドキュメント」をクリックすると右の「アクション」枠にダウンロードボタンがあるのでクリックすると公報全文 PDF (要約、明細書、請求項) が取得できる。ベトナム公報の大半が画像 PDF だが、その公報の詳細画面の「書誌」には要約のテキストがあり、「説明」をクリックすると明細書のテキスト、「請求項」をクリックすると請求項のテキストがそれぞれ取得できる。
- ・ ASEAN IP REGISGER からの検索方法は次の通りである。ASEAN IP REGISGER 日本語版²²¹より、ベトナムの「特許」列の青い数字部分をクリックする。上部の「詳細」をクリックすると、詳細検索画面になるので、上部右にある「フィールドの追加/削除」をクリックして、「国」のオプションメニューから「優先権の詳細」にチェックを入れると、画面中央に入力欄が表示される。「JP」と入力し、「検索」をクリックすると検索結果一覧が表示されるので、閲覧したい公報をクリックすると、その公報の詳細画面から要約のテキストが取得できる。また、他にも色々なフィールドを追加することができるので、条件を絞って検索することができる。

パテントファミリーに日本公報が含まれるものに限定してベトナム公報のデータを取得する方法としては以下が挙げられる。

- ・ Patbase 等の商用の特許検索・特許調査ツールを使用し、パテントファミリーに日本公報を含むベトナム公報の番号リストを出力し、特定されたベトナム公報の番号を使用して IP VIET NAM の検索ウェブサイトから公報データを取得する。
- ・ 詳細検索のメニューから優先権情報に「JP」の文字列を含むベトナム公報を絞り込むことができるため、IP VIET NAM の検索ウェブサイトからベトナム公報とファミリー関係にある日本公報の両方の番号を把握することができる。(その場合の検索条件は次の通りである。IP VIET NAM 日本語版の画面左の詳細検索「国」のオプションメニューから「優先権の詳細」、「年月日」のオプションメニューから「登録日」、さらに「雑用」の右の「その他」をクリックし「ステータス」にチェックを入れると、画面中央にチェックを入れた 3 つの条件の入力欄が表示される。「優先権の詳細」に「JP」と入力、「登録日」欄をクリックすると下に青字でカレンダー

²¹⁹ <https://ipvietnam.gov.vn/en/web/english/home>

²²⁰ [http://wipopublish.ipvietnam.gov.vn/wopublish-search/public/patents?13&lang=jp&query=*.:](http://wipopublish.ipvietnam.gov.vn/wopublish-search/public/patents?13&lang=jp&query=*.)

²²¹ <https://asean-ipregister.wipo.net/wopublish-search/public/home?13&lang=jp>

ーが出てくるので「期間」をクリックし、カレンダー左側の月の「日」の上をクリックすると月が遡るので、例えば2021年7月1日を選択、また、カレンダーの右側の月の「土」の上をクリックすると月が進むので、例えば2023年6月30日を選択して下の「適用」をクリックすると、「登録日」欄には「2021.07.01 TO 2023.06.30」と入力され、2021年7月1日から2023年6月30日の期間に登録された公報の一覧が表示される（日付は手動で変えられるので、適当な期間を選択してから変更してもよい。）。「ステータス」欄は「Published」と入力し、右の「検索」をクリックすると上記3つの条件すべて当てはまった検索結果一覧が表示される。また、検索結果一覧は「登録日」順で並び替えることもできる。

- ・ 調査時の2024年3月時点において、公報全文PDFおよびテキストの両方が取得できる確率が高い期間は2021年7月1日から2023年6月30日の期間で、上記検索条件での検索結果は2033件であった。）

<特記事項>

- ・ 工業所有権情報・研修館（INPIT）のウェブサイトである「新興国等知財情報データバンク」の「ベトナムにおける特許の調べ方（IP VIET NAM ウェブサイト）」²²²に記載された手順を参照するとよい。
- ・ 登録日の期間に関わらず、公報全文PDFは取得できないが、要約、明細書、請求項のテキストは取得できる公報もある。例えば、「優先権の詳細」と「ステータス」を上記検索条件で絞り込んだうえで、登録日が調査時から約20年前の公報でも要約、明細書、請求項のテキストが取得できるものも多くあった。ただし、調査時では登録日が直近8か月以内の公報はPDFおよびテキストともに取得できなかった。テキストを取得する際の注意点として、他言語の公報にも共通するが、公報に記載されていないと思われる番号やHTMLタグが入っていたり、文字化けや請求項1、請求項2等の項番が消えている公報もあるので使用する際は確認が必要と思われる。

5.2.7. タイ語

タイ王国

タイ知的財産局の検索ウェブサイト等からタイ公報を検索して公報データを取得することができる。

<手順>

- ・ タイ知的財産局（Department of Intellectual Property : DIP）の英語版検索ウェブサイト²²³にアクセスし、所定のメニューからデータベースのウェブページ²²⁴に

²²² <https://www.globalipdb.inpit.go.jp/etc/37558/>

²²³ <https://www.ipthailand.go.th/th/home.html>

²²⁴ <https://search.ipthailand.go.th/>

移動する。

- 表示された簡易検索画面の入力欄に検索対象となる出願番号を入力して検索する（出願番号の例：1401002907）。
- 該当公報が表示された場合はクリックしてその公報の詳細画面を開き、画面右下にある青字部分をクリックして公報データを取得する。なお、要約含む書誌、明細書、請求項でファイルが分かれている。
- ASEAN IP REGISGER からの検索方法は次の通りである。ASEAN IP REGISGER 日本語版²²⁵より、タイの「特許」列の青い数字部分をクリックする。上部の「詳細」をクリックすると、詳細検索画面になるので、上部右にある「フィールドの追加/削除」をクリックして、「国」のオプションメニューから「優先権の詳細」にチェックを入れると、画面中央に入力欄が表示される。「JP」と入力し、「検索」をクリックすると検索結果一覧が表示されるので、閲覧したい公報をクリックすると、その公報の詳細画面から要約のテキストが取得できるが、タイ公報では請求項のテキストも取得できるものが多くある。また、他にも色々なフィールドを追加することができるので、条件を絞って検索することができる。

パテントファミリーに日本公報が含まれるものに限定してタイ公報のデータを取得する方法としては以下が挙げられる。

- Patbase 等の商用の特許検索・特許調査ツールを使用し、パテントファミリーに日本公報を含むタイ公報の番号リストを出力し、特定されたタイ公報の番号を使用して DIP の検索ウェブサイトから公報データを取得する。
- DIP のデータベースのページにおいて、ウェブブラウザの機械翻訳機能を使用して表示を日本語に変換し、「詳細検索」のメニューから優先権主張番号が日本の出願番号であるものに限定して公報検索を行って公報データを取得する。

<特記事項>

- 工業所有権情報・研修館（INPIT）のウェブサイトである「新興国等知財情報データバンク」の「タイにおける特許公報へのアクセス方法」²²⁶に記載された手順を参照するとよい。
- タイ公報の大半が画像 PDF であるため、ユーパス作成の際には OCR ツールを使用して公報 PDF を OCR 化する必要がある。（OCR ツールにかける前に公報の余白部分をトリミングして、頁番号や行番号等の不要な文字列を除外するとよい。）
- OCR 処理後のテキストでは特に英数字の部分で文字化けになるものが多い。

5.2.8. アラビア語

アラビア語を公用語とする国・地域は複数あるが、ここではサウジアラビアとエジプト

²²⁵ <https://asean-ipregister.wipo.net/wopublish-search/public/home?13&lang=jp>

²²⁶ <https://www.globalipdb.inpit.go.jp/etc/8249/>

を調査した。

サウジアラビア王国

サウジアラビア知的財産局の検索ウェブサイト等からサウジアラビア公報を検索して公報データを取得することができる。

<手順>

- ・ WIPO PUBLISH のサウジアラビア版検索ウェブサイト²²⁷からサウジアラビア特許を所定の条件で検索することができる。
- ・ 「ベーシック」をクリックするとフリーワード検索が可能。
- ・ 「詳細」をクリックすると書誌事項での絞り込みが可能。例えば、詳細ページの「フィールドの追加/削除」をクリックし、「優先権の詳細」にチェックを入れて追加し、その欄に「JP」と入力して「検索」をクリックすると優先国が日本のものに絞り込むことができる。
- ・ 検索結果一覧が表示されるので、閲覧したい公報をクリックし、その公報の詳細ページに移ったら、「ドキュメント」をクリックする。
- ・ 表示された画面の「アクション」欄にダウンロードを意味するアイコンがあり、「ドキュメントのダウンロード」の表示をクリックすることで公報全文 PDF（要約、明細書、請求項）を取得することができる。サウジアラビア公報の大半が画像 PDF だが、その公報の詳細画面の「書誌」には要約のテキストがあり、「説明」をクリックすると明細書のテキスト、「請求項」をクリックすると請求項のテキストがそれぞれ取得できる。

エジプト・アラブ共和国

エジプト特許庁の検索ウェブサイト等からエジプト公報の検索は可能であるものの、公報データの取得はできないものが多い。

<手順>

- ・ エジプト特許庁の英語版検索ウェブサイト（Search in Egyptian Patent Office Database (EGPO)）²²⁸にアクセスし、所定の条件で公報を検索することができる。
- ・ 「Priority Country」欄の選択肢のアラビア語を確認するためにウェブブラウザの機械翻訳機能を利用して表示を日本語に変換したら「日本」を選択し、「Search」をクリックして検索を行う。
- ・ 検索結果として出願番号の一覧が表示されるので、閲覧したい公報の出願番号をクリックして詳細画面に入る。

パテントファミリーに日本公報が含まれるものに限定してアラビア語の公報データを取得する方法としては以下が挙げられる。

²²⁷ <https://ipsearch.saip.gov.sa/wopublish-search/public/patents?2&lang=jp>

²²⁸ <http://www.egypt.gov.eg/Search/Default.aspx?lang=en>

- ・ Patbase 等の商用の特許検索・特許調査ツールを使用し、パテントファミリーに日本公報を含むアラビア語の各国公報の番号リストを出力し、特定されたアラビア語の各国公報の番号を使用して各特許検索ウェブサイトから公報データを取得する。
- ・ 前述した WIPO PUBLISH の英語版検索ウェブサイト²²⁹から、「ADVANCED SEARCH」の「+/- Fields」をクリックして「TEXT FIELDS」タブの選択肢の中から「Priority Details」にチェックを入れ、その欄に「JP」と入力し「Search」をクリックして優先国が日本である公報に絞り込む。検索結果一覧が表示されるので、閲覧したい公報をクリックし、その公報の詳細ページに入る。詳細ページからアラビア語の要約のテキストを取得し、「Priority Details」欄に表示された日本公報の番号を把握する。

<特記事項>

- ・ 詳細画面からアラビア語と英語両方の要約はテキストで取得できたが、公報全文の記載はなく、公報全文 PDF も取得できなかった。
- ・ アラビア語の要約は HTML タグが入っているものも多いため、コーパス作成の際にはタグ部分を削除する必要がある。
- ・ PATENTSCOPE の Simple Search ページ²³⁰や、Espacenet の Advanced search ページ²³¹等からもアラビア語の公報が取得できるものもある。

5.2.9. ヒンディー語

インド共和国

インド特許意匠商標総局の検索ウェブサイト等からインド公報を検索して公報データを取得することができるが、かなり少ないと思われる。

<手順>

- ・ インド特許意匠商標総局の検索ウェブサイト (Indian Patent Advanced Search System: inPASS)²³²にアクセスし、所定の条件で公報を検索することができる。なお、本サイトは英語版のみ提供されていると思われる。
- ・ 例えば、「Please Enter Abstract」欄に「device」と入力し、下部にあるキャプチャ認証文字を入力の上、「Search」をクリックすると検索結果一覧が表示される。閲覧したい公報の「Application Number」をクリックするとその公報の詳細画面が表示され、要約、明細書、請求項（すべて英語）が表示され、テキストを取得する

²²⁹ <http://www.egypt.gov.eg/wopublish-search/public/patents;jsessionid=88779185885ECFDE5D2BF4C4C2A1BD26?0&lang=en>

²³⁰ <https://patentscope2.wipo.int/search/en/search.jsf>

²³¹ https://worldwide.espacenet.com/advancedSearch?locale=en_EP

²³² <https://iprsearch.ipindia.gov.in/PublicSearch/>

ことができる。また、下部の「View Application Status」をクリックし、次の画面の下部にある「View Documents」をクリックすると、要約、明細書、請求項（すべて英語）の PDF データが取得できる。

- ・「Please Enter Abstract」欄にヒンディー語で入力すると、検索結果一覧に表示される「Title（「発明の名称」の意）」欄は英語で表示されるが、その公報の詳細画面に入ると要約のテキストがヒンディー語で表示され、要約、明細書、請求項の PDF データもヒンディー語のものが取得できる。

<特記事項>

- ・インド特許意匠商標総局の検索ウェブサイトからヒンディー語の公報の検索は可能ではあるが、大半が英語の公報であり、ヒンディー語公報のデータはかなり少ない。例えば、「Please Enter Abstract」欄に英語で「device」と入力して表示された検索結果は 39,470 件だったのに対し、ヒンディー語で「उपकरण（「装置」、「device」の意）」と入力して表示された検索結果はたった 1 件だったので、ヒンディー語公報の取得は難しいと思われる。

5.2.10. トルコ語

トルコ共和国

トルコ特許商標庁の検索ウェブサイト等からトルコ公報を検索して公報データを取得することができる。

<手順>

- ・トルコ特許商標庁（Patent Search turkpatent: TÜRKPATENT）の英語版検索ウェブサイト²³³にアクセスし、所定の条件で公報を検索することができる。
- ・キーワードで検索する場合、「PATENT RESEARCH」タグの「Name of Invention」から、例えば「bilgisayar（「コンピュータ」の意）」と入力し、「Do QUERY」をクリックすると検索結果一覧が表示される（トルコ語で入力しないと検索に引っかからないので注意が必要）。
- ・閲覧したい公報の「DETAILS」をクリックすると、その公報の詳細ページに移る。
- ・「DOCUMENTS」をクリックすると透明テキスト付きの公報全文 PDF（要約、明細書、請求項）をダウンロードすることができる。

パテントファミリーに日本公報が含まれるものに限定してトルコ公報のデータを取得する方法としては以下が挙げられる。

- ・Patbase 等の商用の特許検索・特許調査ツールを使用し、パテントファミリーに日本公報を含むトルコ公報の番号リストを出力し、特定されたトルコ公報の番号を使用して TÜRKPATENT の検索ウェブサイトから公報データを取得する。

²³³ <https://www.turkpatent.gov.tr/en/research?form=patent>

<特記事項>

- ・ TÜRKPATENT のウェブサイトから優先権主張番号が日本の出願番号であるものに限って公報を検索することはできなかった。
- ・ トルコ公報の大半が透明テキスト付き PDF であるためテキストの取得が可能ではあるが、Word 等のファイルにコピーしたときに改行がなくなるので、コーパス作成の前には形式の修正等の作業が必要となる。

5.3. 多言語 OCR ツール

特許情報の機械翻訳における多言語対応において、特許情報のテキスト化はそのスタート地点にあたる工程である。特許情報のテキスト化の現状は、3パターンあり、①特許情報の画像データ化（PDF ほか）で、既に行われているケースが多い、②テキスト化データで、言語により、画像データに対してテキスト化されているケースと、されていないケースがあり、テキスト化されているケースでは、テキストの精度にはバラツキが生じていると推測される。③画像データ内にテキストデータが埋め込まれているケースもある。

テキスト化未対応のケースは、コーパス作成や機械翻訳を行う為には多言語 OCR 等によるテキスト化が必要となる。このため、多言語対応可能な OCR システムの調査を行った。

本章で述べる多言語対応可能な OCR システムの調査にあたり、そのシステムの対応可能なデータ形式・機能・サポート状況等を確認した。さらにテキストの読み取り精度について検証を実施した。

5.3.1. 公開情報調査・ヒアリング調査

調査を実施するにあたり、多言語対応可能な OCR システムベンダーとして計7社を選定した。

しかし、選定した7社中3社は、多言語 OCR 対応サービスの停止や本調査の対象言語外などで、調査対象から除外した。

したがって、試作コーパス言語に対応する OCR システムを提供するシステムベンダー4社の公開情報調査を行った。さらに公開情報調査を進める中で発生した不明点や疑問点をシステムベンダーにオンライン・対面を織り交ぜて、直接ヒアリング調査を行い解消した。

5.3.2. 多言語 OCR 調査項目

多言語 OCR 対応システムの公開情報調査及びヒアリング調査の調査項目は、表 42 とした。

表 42 多言語 OCR 調査項目

| 項目 | 調査内容 |
|-------------------|--------------------|
| 1) システムベンダー名 | システム提供会社名 |
| 2) システム名称 | システムの名称 |
| 3) 開発拠点 | システム開発の拠点（国） |
| 4) 調査言語 | 対応可能な言語 |
| 5) 入力ファイル形式 | 入力可能なデータ形式 |
| 6) 出力ファイル形式 | 出力可能なデータ形式 |
| 7) オンプレ | オンプレミス対応可否 |
| 8) クラウド | クラウド対応可否 |
| 9) セキュリティ対応 | サポートされているセキュリティの詳細 |
| 10) API連携 | API連携状況 |
| 11) 手書きor活字 | 手書き文字・活字の読取対応状況 |
| 12) 定型資料or非定型資料対応 | 定型資料・不定型資料の読込対応可否 |
| 13) テンプレート設定可否 | テンプレート設定可否 |
| 14) システム内修正の可否 | OCR後の情報修正可否 |
| 15) システム内画像補正機能 | 画像補正機能の有無 |
| 16) AI活用 or AI非活用 | AI活用状況 |
| 17) AI-OCR再学習取扱い | AI-OCR再学習権利の取扱い |
| 18) 推奨利用環境 | 対応可能なOS |
| 19) ユーザーサポート | ユーザーサポートの対応詳細 |

5.3.3. 公開情報調査・ヒアリング調査結果

多言語 OCR 対応システムの公開情報調査及びヒアリング調査の調査結果を表 43 として整理した。

表 43 多言語 OCR 公開情報調査&ヒアリング調査結果一覧

| システムベンダー名 | | | A社 | B社 | C社 | D社 | |
|----------------------------|---------|--|---|---|---|-------|---|
| システム名称 | | | W | X | Y | Z | |
| 開発拠点 | | | 米国 | 日本 | 日本 | 米国 | |
| 調査言語 | スペイン語 | Spanish | ラテン文字 | ○ | ○ | × | ○ |
| | ポルトガル語 | Portuguese | ラテン文字 | ○ | ○ | × | ○ |
| | イタリア語 | Italian | ラテン文字 | ○ | ○ | × | ○ |
| | インドネシア語 | Indonesian | ラテン文字 | ○ | ○ | × (※) | ○ |
| | マレー語 | Malay | ラテン文字 | ○ | ○ | × | ○ |
| | ベトナム語 | Vietnamese | ラテン文字 | ○ | ○ | ○ | ○ |
| | タイ語 | Thai | タイ文字 | ○ | ○ | ○ | ○ |
| | アラビア語 | Arabic | アラビア文字 | ○ | × | × | ○ |
| | ヒンディー語 | Hindi | デーヴァナーガリー文字 | ○ | × | × | × |
| | トルコ語 | Turkish | ラテン文字 | ○ | × | × | ○ |
| 入力ファイル形式 | | JPEG PNG8 PNG24 GIF アニメーション GIF (最初のフレームのみ) BMP WEBP RAW ICO PDF TIFF | BMP TIFF マルチページTIFF (非圧縮, G3/G4圧縮, LZW圧縮, PackBits) PNG JPEG PDF (画像のみ) | PDF JPEG PNG TIFF | TIFF, Multipage TIFF (非圧縮, CCITT3, CCITT3FAX, CCITT4, PackBits) JPEG (ZIP, LZW (8/24bits)) JPEG 2000 JBIG2 BMP GIF PNG WDP XPS PCX, DCX PDF DjVu DOC, DOCX, ODT XLS, XLSX, ODS PPT, PPTX, ODP TXT, HTML, HTM, RTF | | |
| 出力ファイル形式 (特に透明テキスト付PDF) | | テキストにて出力 | テキスト Unicodeテキスト RTF DOCX XLSX PPTX CSV HTML PDF SON | CSV PDF | PDF (A-1a, A-1b, -2a, A-2b, A-2u, -3a, A-3b, A-3u, E, UA) RTF DOC, DOCX XLS, XLSX PPTX TXT CSV HTML TIFF JPEG JPEG 2000 PNG EPUB XML Alto XML JSON | | |
| オンプレ | | × | ○ | ○ | ○ | | |
| クラウド | | ○ | × | ○ | × | | |
| セキュリティ対応 | | 規約に準ずる | オンプレのため非該当 | (有償オプション) IP制限 VPN Direct Connect Private Link | オンプレのため非該当 | | |
| API連携 | | ○ | ○ | ○ | ○ | | |
| 手書きor活字 | | 双方対応 | 活字 | 双方対応 | 活字 | | |
| 定型資料or非定型資料対応 | | 双方対応 | 非定型 | 双方対応 | 全文OCRのみ対応 | | |
| テンプレート設定可否 | | × | ○ | ○ | ○ | | |
| システム内修正の可否 | | × | × | ○ | ○ | | |
| システム内画像補正機能 | | × | ○ | ○ | ○ | | |
| AI活用 or AI非活用 | | AI活用 (画像認識技術として) | AI非活用 | AI活用 | AI活用 | | |
| AI-OCR再学習取扱い | | 学習対象外 | AI非活用のため該当なし | 学習機能ON/OFF制御可能 個人情報含むデータは学習対象外 | 学習対象外 | | |
| 推奨利用環境 | | 依存なし | Microsoft Windows 11 Microsoft Windows 10 (64bit) Windows Server 2022 Windows Server2019 2016, CloudAPI | Microsoft Windows 10 Microsoft Windows 11 macOS 10.12 (Sierra) 以上 | Microsoft® Windows® 11 Windows® 10 Windows® 8.1 Windows Server® 2022 Windows Server® 2019 Windows Server® 2016 Windows Server® 2012 R2 Windows Server® 2012 Microsoft Azure Virtual Machines | | |
| ユーザーサポート | | ○ | ○ | ○ | ○ | | |

※C社のインドネシア語はアルファベット表記につき英語版で代用

5.3.4. 公開情報調査・ヒアリング調査考察

全体の考察は以下の通りである。

- ・入力、出力形式においては各システムにおいて形式数やバリエーションに差がある。

- ・ オンプレ、クラウド対応については、片方のみ対応と双方対応可能のシステムがある。
- ・ テキスト化後の修正については、システム内での修正機能が備わっているものと、システム外で修正するものに分かれる。

さらにヒアリング調査による特筆すべき情報は以下の通りである。各 OCR システムは以下の通りである。

- ・ A 社（システム名：W）

多言語対応が強み（今回対象とした 10 言語すべて対応可能）。処理方法は、ドラッグ & ドロップにより画像データを読み込み、OCR 処理してテキスト化。2,000 ページまで PDF、TIFF ファイルの受け入れが可能。

- ・ B 社（システム名：X）

調査言語 10 言語のうち、7 言語に対応可能。自動的に図の置換や表・下線の判断する機能がある。さらに精度向上のため、文章優先や図表優先で読み取る設定が可能。

- ・ C 社（システム名：Y）

調査言語 10 言語のうち、3 言語に対応可能。大量の帳票を得意とする OCR システム（定型資料に強い）で、AI-OCR を活用したシステム。

- ・ D 社（システム名：Z）

調査言語 10 言語のうち、ヒンディー語を除く 9 言語に対応可能。大量の文書に対応した OCR システムで、複雑な帳票の読取にも長ける。

5.3.5. 試作コーパス言語の OCR 精度検証

選定した 4 社のシステムを使用して試作コーパス言語（インドネシア語・ベトナム語・タイ語）に対して、実際に OCR 処理を行い、精度検証を実施した。OCR 処理において得手不得手に分かれやすい各種組版バリエーションと全体に対して、精度検証を行った。

5.3.6. 精度検証設定

精度検証データは、試作コーパス言語（3 言語）の特許情報の中から組版バリエーションを加味して、各言語 4 頁（4 頁／言語× 3 言語＝計 12 頁）を選定した。

選定した資料は、インドネシア語（出願番号：P00202203012、P00202200077、P00202200065）、タイ語（出願番号：TH20191008103、TH20211008078）、ベトナム語（2015 年 9 月発表分の公報一覧：330A）である。

OCR 処理において各組版バリエーション（①本文 ②表組 ③下線 ④フローチャート ⑤化学式 ⑥数式 ⑦グラフ）と全体に対する精度検証を行った。いずれも OCR 処理のみで検証することで、同じ条件下での精度結果を得た。組版バリエーションとは以下の通り

である。

- ①本文：明細書内の主たる内容を成す基本的な組版
- ②表組：表形式に作成された組版

58

nikotin:resin: 1,19:1 (1,19)
komposisi nikotin-resin yang diperoleh: 22,8
proses adalah 20 menit.

5

①本文

Contoh 2H - Premiks nikotin VIII - Resin

Suatu pengaduk planetary Bear Varimixer 60 liter diisi dengan air, dan nikotin ditimbang dan ditambahkan. Pengaduk diaduk dengan kecepatan rendah selama 1 menit pada suhu lingkungan. Kemudian resin pertukaran ion

10 Amberlite ® IRP64 ditimbang dan dimasukkan ke dalam pengaduk. Pengaduk ditutup, diaduk dengan kecepatan tinggi selama 5 menit, dibuka dan dikikis, jika diperlukan. Terakhir, pengaduk diaduk lagi selama 5 menit

15 dengan kecepatan tinggi.

Dengan demikian, suatu campuran dari pertukaran ion nikotin dan catresin dihasilkan dari konstituen yang ditunjukkan dalam tabel 4E di bawah ini.

20

②表組

an yang digunakan untuk pembuatan premiks nikotin VIII.

| Konstituen | Jumlah (kg) | Jumlah (%) |
|---------------------|-------------|------------|
| Nikotin | 5,15 | 39,8 |
| Air | 2,80 | 21,6 |
| Resin | 4,32 | 33,4 |
| Berat kacang polong | 0,67 | 5,2 |
| Total | 12,94 | 100,0 |

Perbandingan nikotin:resin: 1,19:1 (1,19)
% air dalam komposisi nikotin-resin yang diperoleh: 21,6
25 Total waktu proses adalah 20 menit.

図 32 組版バリエーション例 (①本文②表組)

- ③下線：文章の直下に引く線で、強調に用いる

9

dan (koefisien dekorrelasi nyata) dapat dihitung dari elemen matriks kovarians sisi-tengah (2x2). Matriks ini dihitung melalui serangkaian pita analisis $1 \leq b \leq B$.

$$\Sigma_{MS}(b) = \begin{bmatrix} R_{MM}(b) & R_{MS}(b) \\ R_{SM}(b) & R_{SS}(b) \end{bmatrix}$$

5 persamaan 1

$$a(b) = \frac{R_{SS}(b)}{R_{MM}(b)}, \beta(b) = \frac{R_{SM}(b)}{\sqrt{R_{MM}(b)}} - a(b)^2$$

dimana istilah dalam bentuk $R_{ij}(b) = X(i)Y(j)^*$ adalah kovarians sinyal X dan Y yang dihitung pada pita frekuensi b. Superskrip * menunjukkan transpos konjugat kompleks dari nilai. Contoh di atas adalah matriks kovarians 2 x 2. Dalam berbagai implementasi, sejumlah saluran masukan dimungkinkan (misalnya, sistem empat saluran dengan matriks kovarians 4 x 4). Contoh di atas adalah untuk sinyal tengah (M) dan samping (S), tetapi orang yang ahli dalam bidang ini dapat memperkirakan di atas untuk jenis sinyal lainnya.

Fungsi Pemulusan dan Faktor Lupa

20 Umumnya, matriks yang dihaluskan menggunakan filter lolos rendah yang memenuhi persyaratan pemulusan. Dalam kasus algoritma pemulusan, persyaratan pemulusan sedemikian rupa sehingga perkiraan sebelumnya digunakan untuk secara artifisial meningkatkan jumlah sampel frekuensi (nampan) yang digunakan untuk menghasilkan perkiraan matriks kovarians saat ini. Dalam beberapa perwujudan, menghitung matriks yang dihaluskan dari matriks kovarians masukan pada sekuens kerangka menggunakan filter lolos rendah regresif otomatis orde pertama yang menggunakan jumlah bobot dari estimasi nilai matriks kerangka sebelumnya dan sekarang:

③下線

図 33 組版バリエーション例 (③下線)

④フローチャート：プロセス・システムやプログラム処理等を示す流れ図

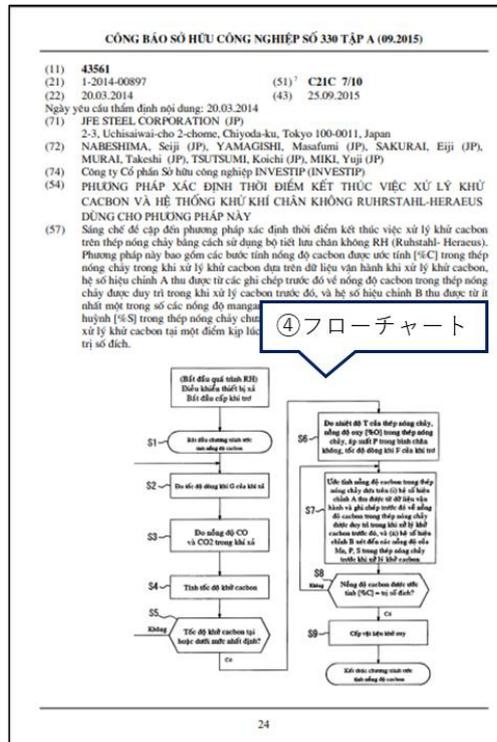


図 34 組版バリエーション例 (④フローチャート)

⑤化学式：化学物質の成り立ちを元素記号と数字で表したもの

4

Aspek lain dari invensi ini berkaitan dengan senyawa formula umum (II)



⑤化学式

(II)

5 di mana

R⁴ adalah C₁-C₁-alkil atau gugus *- [2-O]_n-CH=CH₂, di mana tanda bintang menunjukkan titik pelekatan ke atom oksigen, 2 adalah C₁-C₁-alkilena dan n adalah 0, 1, 2, 3, 4 atau 5, stereoisomernya atau campuran stereoisomernya.

10 Aspek lain dari invensi ini berkaitan dengan senyawa formula umum (III)



(III)

stereoisomernya atau campuran stereoisomernya.

15 Aspek lainnya lagi dari invensi ini berkaitan dengan proses untuk membuat senyawa formula umum (I), atau campuran dua atau lebih senyawa yang berbeda dari formula umum (I) atau stereoisomer dari senyawa (I) atau campuran dua atau lebih stereoisomer dari senyawa (I) atau campuran dua atau lebih

20 stereoisomer dari dua atau lebih senyawa berbeda (I), yang prosesnya meliputi:

(i) menyediakan 2,5,6-trimetilseks-2-en-1-ol (IV),



(IV)

図 35 組版バリエーション例 (⑤化学式)

⑥数式：数字・記号を結合した文字列

9

dan (koefisien dekorrelasi nyata) dapat dihitung dari elemen matriks kovarians sisi-tengah (2x2). Matriks ini dihitung melalui serangkaian pita analisis $1 \leq b \leq B$.

$$\Sigma_{MS}(b) = \begin{bmatrix} R_{MM}(b) & R_{MS}(b) \\ R_{SM}(b) & R_{SS}(b) \end{bmatrix}$$

⑥数式

5 persamaan 1

$$\alpha(b) = \frac{R_{SM}(b)}{R_{SS}(b)}, \beta(b) = \sqrt{\frac{R_{MM}(b)}{R_{SS}(b)} - \alpha(b)^2}$$

dimana istilah dalam bentuk $R_{XY}(b) = X(b)Y(b)^*$ adalah kovarians sinyal X dan Y yang dihitung pada pita

10 frekuensi b. Superskrip * menunjukkan transpos konjugat kompleks dari nilai. Contoh di atas adalah matriks kovarians 2 x 2. Dalam berbagai implementasi, sejumlah saluran masukan dimungkinkan (misalnya, sistem empat saluran dengan matriks kovarians 4 x 4). Contoh di atas

15 adalah untuk sinyal tengah (M) dan samping (S), tetapi orang yang ahli dalam bidang ini dapat memperkirakan di atas untuk jenis sinyal lainnya.

Fungsi Pemulusan dan Faktor Lupa

20 Umumnya, matriks yang dihaluskan dapat dihitung menggunakan filter lolos rendah yang dirancang untuk memenuhi persyaratan pemulusan. Dalam kasus algoritma pemulusan, persyaratan pemulusan sedemikian rupa sehingga perkiraan sebelumnya digunakan untuk secara

25 artifisial meningkatkan jumlah sampel frekuensi (nampan) yang digunakan untuk menghasilkan perkiraan matriks kovarians saat ini. Dalam beberapa perwujudan, menghitung matriks yang dihaluskan dari matriks kovarians masukan pada sekuens kerangka menggunakan

30 filter lolos rendah regresif otomatis orde pertama yang menggunakan jumlah bobot dari estimasi nilai matriks kerangka sebelumnya dan sekarang:

図 36 組版バリエーション例 (⑥数式)

⑦グラフ：棒グラフや折れ線グラフなど、情報を整理して可視化した図表

CÔNG BÁO SỞ HỮU CÔNG NGHIỆP SỐ 330 TẬP A (09.2015)

(11) 43570
 (21) 1-2014-03672 (51)⁷ A61K 39/02, 39/12, 39/295
 (22) 03.04.2013 (43) 25.09.2015
 (86) PCT/US2013/035083 03.04.2013 (87) WO2013/152081 10.10.2013
 (30) 61/620.165 04.04.2012 US
 Ngày yêu cầu thẩm định nội dung: 03.11.2014
 (71) ZOETIS LLC (US)
 100 Campus Drive, Florham Park, New Jersey 07932, United States of America
 (72) Jeffrey E. GALVIN (US), Gregory P. NITZEL (US), John Keith GARRETT (US),
 James R. KULAWIK II (US), Tracy L. RICKER (US), Megan Marie SMUTZER (IN)
 (74) Công ty TNHH T&T INVENMARK Sở hữu trí tuệ Quốc tế (T&T INVENMARK CO.,
 LTD.)
 (54) CHẾ PHẨM MIỄN DỊCH CHỨA PHẦN HÒA TAN CỦA CHẾ PHẨM TẾ BÀO
 MYCOPLASMA HYOPNEUMONIAE TOÀN PHẦN, PHƯƠNG PHÁP BẢO CHẾ
 VÀ KIT CHỨA CHẾ PHẨM MIỄN DỊCH NÀY.
 (57) Sáng chế đề cập đến chế phẩm miễn dịch chứa phần **⑦** M. lyo được theo dõi bằng % tổn thương phổi. Ngoài ra, sá
 bảo chế và kit chứa chế phẩm miễn dịch này. song

Điều trị bằng M. lyo được theo dõi bằng % tổn thương phổi.

| Điều trị | Bình phương trung bình số nhiễm của tổn thương phổi (%) |
|-----------------------|---|
| Phorbol (Compazine) | 11.7 |
| Sốt lục (Sốt lục) | 5.0 |
| Sốt lục (Sốt lục) | 4.7 |
| Sốt lục (Sốt lục) | 4.6 |
| Sốt lục (Sốt lục) | 4.5 |
| Sốt lục (Sốt lục) | 3.7 |
| Dịch sữa Protein A | 3.1 |
| R50 M. lyo | 2.1 |
| Dịch sữa (Dịch sữa) | 1.4 |
| Trùng đơn (Trùng đơn) | 1.2 |
| Trùng đơn (Trùng đơn) | 0.3 |

33

図 37 組版バリエーション例 (⑦グラフ)

精度検証の手順としては、図 38 の通りである。

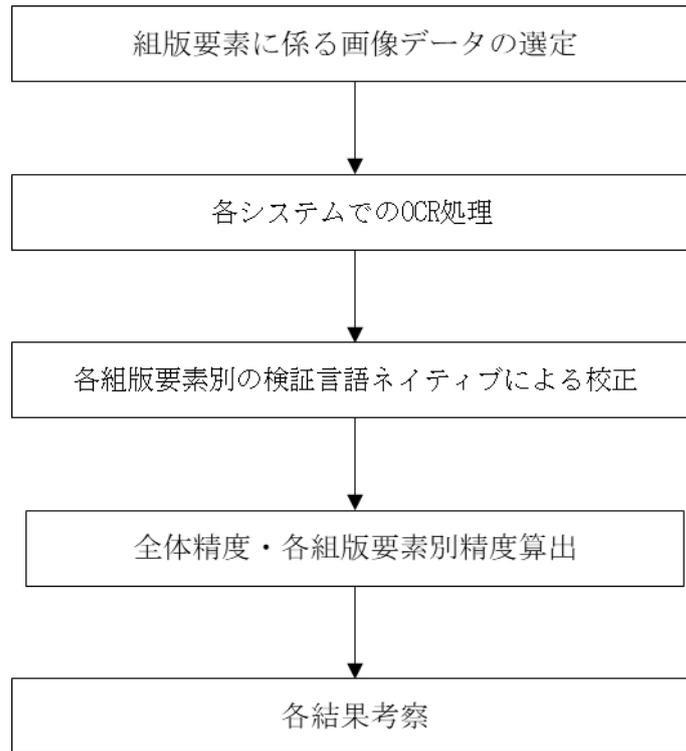


図 38 精度検証手順

検証のポイントとしては、下記の基本的な仕様 4 点で行った。

- ・ 試作コーパス言語：インドネシア語・ベトナム語・タイ語
- ・ 組版バリエーション：
 - ①本文 ②表組 ③下線 ④フローチャート ⑤化学式 ⑥数式 ⑦グラフ
- ・ 頁数：各言語 4 頁（4 頁／言語×3 言語＝計 12 頁）
- ・ OCR 設定：事前設定無く OCR 処理のみ

5.3.7. 精度検証結果

各システムの組版要素別全体精度検証結果は表 44 の通りである。

表 44 各組版要素別の精度検証結果

| 言語名 | インドネシア語 | | | | | | | | タイ語 | | | | | | | | ベトナム語 | | | | | | | | | |
|---|---------|-------|---------|-------|---------|-------|---------|-------|---------|-------|---------|-------|--------|-------|--------|-------|---------|-------|---------|-------|--------|-------|--------|-------|---------|------|
| | A社 | | B社 | | C社 | | D社 | | A社 | | B社 | | C社 | | D社 | | A社 | | B社 | | C社 | | D社 | | | |
| ベンダー名 | W | | X | | Y | | Z | | W | | X | | Y | | Z | | W | | X | | Y | | Z | | | |
| システム名称 | 誤読数 | 全文文字数 | 誤読数 | 全文文字数 | 誤読数 | 全文文字数 | 誤読数 | 全文文字数 | 誤読数 | 全文文字数 | 誤読数 | 全文文字数 | 誤読数 | 全文文字数 | | |
| | 正答率 | | 正答率 | | 正答率 | | 正答率 | | 正答率 | | 正答率 | | 正答率 | | 正答率 | | 正答率 | | 正答率 | | 正答率 | | 正答率 | | | |
| 組版 パ リ エ ー シ ョ ン 種 類 | 表組 | | 1 | 1147 | 545 | 1147 | 7 | 1147 | 0 | 1147 | 1 | 695 | 347 | 695 | 7 | 695 | 26 | 695 | 0 | 782 | 24 | 782 | 1 | 782 | 0 | 782 |
| | | | 99.91% | | 52.48% | | 99.39% | | 100.00% | | 99.86% | | 50.07% | | 98.99% | | 96.26% | | 100.00% | | 96.93% | | 99.87% | | 100.00% | |
| | 下線 | | 0 | 28 | 0 | 28 | 0 | 28 | 0 | 28 | 0 | 85 | 76 | 85 | 2 | 85 | 0 | 85 | - | - | - | - | - | - | - | - |
| | | | 100.00% | | 100.00% | | 100.00% | | 100.00% | | 100.00% | | 10.59% | | 97.65% | | 100.00% | | - | | - | | - | | - | |
| | チャート | | 38 | 452 | 452 | 452 | 452 | 452 | 180 | 452 | 140 | 1410 | 1406 | 1410 | 686 | 1410 | 703 | 1410 | 8 | 675 | 266 | 675 | 97 | 675 | 77 | 675 |
| | | | 91.59% | | 0.00% | | 0.00% | | 60.18% | | 90.07% | | 0.28% | | 51.35% | | 50.14% | | 98.81% | | 60.59% | | 85.63% | | 88.59% | |
| | 化学式 | | 4 | 19 | 16 | 19 | 15 | 19 | 7 | 19 | 29 | 78 | 57 | 78 | 48 | 78 | 78 | 78 | 3 | 17 | 2 | 17 | 13 | 17 | 9 | 17 |
| | | | 78.95% | | 15.79% | | 21.05% | | 63.16% | | 62.82% | | 26.92% | | 38.46% | | 0.00% | | 82.35% | | 88.24% | | 23.53% | | 47.06% | |
| | 数式 | | 18 | 70 | 8 | 70 | 43 | 70 | 68 | 70 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| | | | 74.29% | | 88.57% | | 38.57% | | 2.86% | | - | | - | | - | | - | | - | | - | | - | | - | |
| | グラフ | | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 30 | 327 | 327 | 327 | 327 | 327 | 255 | 327 |
| | | | - | | - | | - | | - | | - | | - | | - | | - | | 90.83% | | 0.00% | | 0.00% | | 22.02% | |
| | 本文 | | 2 | 2827 | 15 | 2827 | 16 | 2827 | 16 | 2827 | 111 | 2990 | 242 | 2990 | 124 | 2990 | 186 | 2990 | 2 | 3055 | 53 | 3055 | 35 | 3055 | 41 | 3055 |
| | | | 99.93% | | 99.47% | | 99.43% | | 99.43% | | 96.29% | | 91.91% | | 95.85% | | 93.78% | | 99.93% | | 98.27% | | 98.85% | | 98.66% | |
| | 全体 | | 63 | 4543 | 1036 | 4543 | 533 | 4543 | 271 | 4543 | 281 | 5258 | 2128 | 5258 | 867 | 5258 | 993 | 5258 | 43 | 4856 | 672 | 4856 | 473 | 4856 | 382 | 4856 |
| | | | 98.61% | | 77.20% | | 88.27% | | 94.03% | | 94.66% | | 59.53% | | 83.51% | | 81.11% | | 99.11% | | 86.16% | | 90.26% | | 92.13% | |

※「-」は選定した明細書ページに該当する組版が無かった項目に記入

5.3.8. 精度検証結果考察

①全体について

A社「W」は各言語の正答率が高く、言語間のバラツキも少ない。日本語 OCR（約 95%）と比較しても 3 言語合計平均の正答率 97.4% は高い水準である。

- 各システムにおいて、得意な言語・不得意な言語や得意な字形・不得意な字形がある。たとえば、不得意な字形としてベトナム語における補助記号「 \hat{A} 」「 \hat{o} 」「 \hat{U} 」等がある。

②各言語について

- インドネシア語

英語ベースの言語で他の言語と比較して正答率が高い。ただし、今回チャートの読み取り難易度が高かった為、正答率を下げている。これは本文の正答率に鑑みれば、他の言語よりも本文や下線の正答率が高いことから読み取れる。

システム間において、最も正答率が高いのは A 社、次いで D 社である。他のシステムはチャート、化学式の正答率が低く、全体の正答率に影響している。

- タイ語

言語の表記体系が他の2言語と比較して、タイ語独自（他の2言語はラテン文字）であるため正答率が低く、読み取り難易度が高い。

システム間において、最も正答率が高いのはA社、次いでC社である。他のシステムはインドネシア語と同様にチャート、化学式の正答率が低く、全体の正答率に影響している。

- ・ベトナム語

ベトナム語もインドネシア語と同様に表記体系がラテン文字であることから、正答率が高い。ただし、補助記号によりインドネシア語と比較して正答率を下げている。

システム間において、最も正答率が高いのはA社、次いでD社である。他のシステムはグラフの正答率が非常に低く、全体の正答率に影響している。

③組版バリエーションについて

- ・本文、表組、下線については、言語によらず、概ね正答率が高い。これらは基本的な文書形式であることから、OCRの「文字を読み取る」といった基本的なスペックが高いと想定される。
- ・日本語OCRにおいても読み取り難易度の高いチャート、化学式、数式、グラフといった複雑な組版においては正答率の低下やバラツキが生じている。これは、文字と罫線や図、記号等を区別して読み取るといった応用的なスペックの違いによるものと想定される。

5.4. 文分割ツール

本節では各言語の文書を文単位に分割するためのツールについて報告する。なおいくつかのツールは多言語に対応しているため、これらのツールについてはその詳細を後述し、各言語に特化したものを先に順に列挙する。なお実行方法については、コマンドラインとして実行できるもののみ記載し、他のプログラムから `import` する必要があるものなどは省略している。

5.4.1. スペイン語

CoreNLP（後述）

spaCy（後述）

5.4.2. ポルトガル語

spaCy（後述）

5.4.3. イタリア語

CoreNLP (後述)

5.4.4. インドネシア語

spaCy (後述)

id-sentence-segmenter

- 機能：文分割
- 利用条件：不明（記載なし）
- 入手方法：<https://github.com/yudanta/id-sentence-segmenter>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

5.4.5. マレー語

malaya

- 機能：文分割、単語分割、品詞タグ付け、レンマ化、固有表現認識、句構造解析、依存構造解析、共参照解析など
- 利用条件：MIT
- 入手方法：<https://malaya.readthedocs.io/en/stable/>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

5.4.6. ベトナム語

Underthesea

- 機能：文分割、テキスト標準化、単語分割、品詞タグ付け、チャンキング、依存構造解析、固有表現認識、テキスト分類、感情分析
- 利用条件：GNU General Public License v3.0
- 入手方法：<https://github.com/undertheseanlp/underthesea>
- プログラミング言語：Python
- ツールのコスト：無料

- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

5.4.7. タイ語

thai-segmenter

- 機能：文分割、単語分割、品詞タグ付け
- 利用条件：MIT
- 入手方法：<https://github.com/Querela/thai-segmenter>
- プログラミング言語：Python
- 実行方法：`thai-segmenter sentseg -i input.txt -o output.txt`
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし（そもそもは文分割をするためのツールだが、文分割のために単語分割や品詞タグ付けの情報を使うため、これらも行うとのこと。）

PyThaiNLP

- 機能：文分割、単語分割、品詞タグ付けなど
- 利用条件：Apache License 2.0
- 入手方法：<https://pythainlp.github.io/>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：様々なツールの寄せ集めのようなので、それぞれの機能を使うためには別々のツールをインストールする必要がある。

5.4.8. アラビア語

CoreNLP（後述）

5.4.9. ヒンディー語

Indic NLP Library

- 機能：テキスト標準化、言語判定（インド諸語）、文分割、単語分割
- 利用条件：MIT
- 入手方法：https://github.com/anoopkunchukuttan/indic_nlp_library
- プログラミング言語：Python
- ツールのコスト：無料

- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

5.4.10. トルコ語

TrTokenizer

- 機能：文分割、単語分割
- 利用条件：MIT
- 入手方法：<https://github.com/apdullahyayik/TrTokenizer>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

5.4.11. 多言語対応ツール

CoreNLP

- 対応言語：アラビア語、中国語、英語、フランス語、ドイツ語、ハンガリー語、イタリア語、スペイン語
- 機能：文分割、単語分割、品詞タグ付け、レンマ化、固有表現認識、句構造解析、依存構造解析、共参照解析など
- 利用条件：GNU General Public License v3 or later
- 入手方法：<https://stanfordnlp.github.io/CoreNLP/>
- プログラミング言語：Java
- 実行方法：`java -cp "*" edu.stanford.nlp.pipeline.StanfordCoreNLP -file input.txt`
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：Java 8+（無料）
- ツールの課題：なし

spaCy

- 対応言語：カタルーニャ語、中国語、クロアチア語、デンマーク語、オランダ語、英語、フィンランド語、フランス語、ドイツ語、ギリシャ語、イタリア語、日本語、韓国語、リトアニア語、マケドニア語、ノルウェー語、ポーランド語、ポルトガル語、ルーマニア語、ロシア語、スロベニア語、スペイン語、スウェーデン語、ウクライナ語（これらは多くの機能が備わっている言語で、他にも多くの言語に対応しているが、使える機能が少ない）
- 機能：文分割、単語分割、品詞タグ付け、レンマ化、固有表現認識、句構造解析、依存構造解析、共参照解析など
- 利用条件：MIT

- 入手方法 : <https://spacy.io/>
- プログラミング言語 : Python
- ツールのコスト : 無料 (高精度・高機能な有料版もあり)
- ツールと依存関係になる他のソフトおよびそのコスト : Cython (無料)
- ツールの課題 : 言語によって対応している機能が少ないものもある

5.5. 形態素解析 (単語分割) ツール

本節では各言語の文を形態素 (単語) 単位に分割するためのツールについて報告する。なおいくつかのツールは多言語に対応しているため、これらのツールについてはその詳細を後述する。

5.5.1. スペイン語

CoreNLP (後述)

spaCy (後述)

5.5.2. ポルトガル語

spaCy (後述)

5.5.3. イタリア語

CoreNLP (後述)

5.5.4. インドネシア語

nlp-id

- 機能 : 単語分割、レンマ化、品詞タグ付け
- 利用条件 : MIT
- 入手方法 : <https://pypi.org/project/nlp-id/>
- プログラミング言語 : Python
- ツールのコスト : 無料
- ツールと依存関係になる他のソフトおよびそのコスト : なし
- ツールの課題 : なし

5.5.5. マレー語

malaya

- 機能：文分割、単語分割、品詞タグ付け、レンマ化、固有表現認識、句構造解析、依存構造解析、共参照解析など
- 利用条件：MIT
- 入手方法：<https://malaya.readthedocs.io/en/stable/>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

5.5.6. ベトナム語

Underthesea

- 機能：文分割、テキスト標準化、単語分割、品詞タグ付け、チャンキング、依存構造解析、固有表現認識、テキスト分類、感情分析
- 利用条件：GNU General Public License v3.0
- 入手方法：<https://github.com/undertheseanlp/underthesea>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

VnCoreNLP

- 機能：単語分割、品詞タグ付け、固有表現認識、依存構造解析
- 利用条件：GNU
- 入手方法：<https://github.com/vncorenlp/VnCoreNLP>
- プログラミング言語：Java か Python ラッパー
- 実行方法：`java -Xmx2g -jar VnCoreNLP-1.2.jar -fin input.txt -fout output.txt -annotators wseg,pos,ner`
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：Java 1.8+ (無料)
- ツールの課題：なし

PyVI

- 機能：単語分割、品詞タグ付け、アクセントの除去/付加
- 利用条件：MIT
- 入手方法：<https://github.com/trungtv/pyvi>
- プログラミング言語：Python
- ツールのコスト：無料

- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：最終更新が2年前なので古い

5.5.7. タイ語

thai-segmenter

- 機能：文分割、単語分割、品詞タグ付け
- 利用条件：MIT
- 入手方法：<https://github.com/Querela/thai-segmenter>
- プログラミング言語：Python
- 実行方法：`thai-segmenter sentseg -i input.txt -o output.txt`
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし（そもそもは文分割をするためのツールだが、文分割のために単語分割や品詞タグ付けの情報を使うため、これらも行うとのこと。）

PyThaiNLP

- 機能：文分割、単語分割、品詞タグ付けなど
- 利用条件：Apache License 2.0
- 入手方法：<https://pythainlp.github.io/>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：様々なツールの寄せ集めのようなので、それぞれの機能を使うためには別々のツールをインストールする必要がある。

5.5.8. アラビア語

CoreNLP（後述）

5.5.9. ヒンディー語

Indic NLP Library

- 機能：テキスト標準化、言語判定（インド諸語）、文分割、単語分割
- 利用条件：MIT
- 入手方法：https://github.com/anoopkunchukuttan/indic_nlp_library
- プログラミング言語：Python
- ツールのコスト：無料

- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

5.5.10. トルコ語

TrTokenizer

- 機能：文分割、単語分割
- 利用条件：MIT
- 入手方法：<https://github.com/apdullahyayik/TrTokenizer>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

5.5.11. 多言語対応ツール

CoreNLP

- 対応言語：アラビア語、中国、英語、フランス語、ドイツ語、ハンガリー語、イタリア語、スペイン語
- 機能：文分割、単語分割、品詞タグ付け、レンマ化、固有表現認識、句構造解析、依存構造解析、共参照解析など
- 利用条件：GNU General Public License v3 or later
- 入手方法：<https://stanfordnlp.github.io/CoreNLP/>
- プログラミング言語：Java
- 実行方法：`java -cp "*" edu.stanford.nlp.pipeline.StanfordCoreNLP -file input.txt`
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：Java 8+（無料）
- ツールの課題：なし

spaCy

- 対応言語：カタルーニャ語、中国語、クロアチア語、デンマーク語、オランダ語、英語、フィンランド語、フランス語、ドイツ語、ギリシャ語、イタリア語、日本語、韓国語、リトアニア語、マケドニア語、ノルウェー語、ポーランド語、ポルトガル語、ルーマニア語、ロシア語、スロベニア語、スペイン語、スウェーデン語、ウクライナ語（これらは多くの機能が備わっている言語で、他にも多くの言語に対応しているが、使える機能が少ない）
- 機能：文分割、単語分割、品詞タグ付け、レンマ化、固有表現認識、句構造解析、依存構造解析、共参照解析など
- 利用条件：MIT

- 入手方法 : <https://spacy.io/>
- プログラミング言語 : Python
- ツールのコスト : 無料 (高精度・高機能な有料版もあり)
- ツールと依存関係になる他のソフトおよびそのコスト : Cython (無料)
- ツールの課題 : 言語によって対応している機能が少ないものもある

5.6. アライメントツール

本節では対訳文アライメントツールについて報告する。対訳文アライメントは大きく分けて I. 文ペアの類似度を計算するステップと II. 類似度を元に文書ペアから対応する文ペアを抽出するステップとからなる。I. については a. 対訳辞書を用いる方法、b. 文をベクトル化する方法、c. 機械翻訳を用いる方法などがある。II. については動的計画法 (Dynamic Programming) や貪欲法 (Greedy Algorithm) などがある。それぞれについて利用可能なツールを以下に示す。ただし、II. についてはこの部分のみを実装したツールは見当たらなかった (I. を実装したツール内で II. も実装されている) ため、II. については割愛する。

I-a. 対訳辞書を用いる方法

- 概要 : 文ペア内に存在する対訳関係にある単語の割合などに基づいて類似度スコアを計算する方法。
- 利点 : 対訳辞書に記載されている情報を用いるため、対応付く単語が多ければ対訳である可能性が高く、精度も良い。
- 課題 : 対訳辞書を用意する必要があり、対訳コーパス構築の精度は対訳辞書の規模と精度に依存する。
- ツール :

Champollion

- 対応言語 : 英語とアラビア語、中国語、ヒンディー語
- 利用条件 : Common Public License
- 入手方法 : <https://github.com/LowResourceLanguages/champollion>
- プログラミング言語 : Perl
- ツールのコスト : 無料
- ツールと依存関係になる他のソフトおよびそのコスト : なし
- ツールの概要 : 対訳辞書がないと使えない。対象言語については小規模な辞書が同梱されている。辞書を入れ替えれば、どの言語でも利用可能。ただし単語分割については自前で行う必要がある。

hunalign

- 利用条件 : GNU LGPL version 2.1 or later
- 入手方法 : <http://mokk.bme.hu/en/resources/hunalign/>

- プログラミング言語：C++
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの概要：対訳辞書を用いたアライメント手法。対訳辞書がない場合、文の長さなどから簡易アライメントを行い、その結果から対訳辞書を自動構築し、再度辞書ベースのアライメントを行う。

I-b. 文をベクトル化する方法

- 概要：事前学習されたモデルを用いて各言語の各文を、言語非依存の共通のベクトル空間にマッピングする。マッピング後の距離を類似度として用いる。
- 利点：対訳辞書が不要で、多くの言語に対応しているツールが多い。
- 課題：文が短い場合や固有名詞が多い場合などはうまくベクトル化できず、誤った文との類似度が高いと判定されることがある。
- ツール：

Bertalign

- 対応言語：カタルーニャ語、中国語、チェコ語、デンマーク語、オランダ語、英語、フィンランド語、フランス語、ドイツ語、ギリシャ語、ハンガリー語、アイスランド語、イタリア語、リトアニア語、ラトビア語、ノルウェー語、ポーランド語、ポルトガル語、ルーマニア語、ロシア語、スロバキア語、スロベニア語、スペイン語、スウェーデン語、トルコ語
- 利用条件：GNU General Public License v3.0
- 入手方法：<https://github.com/bfsujason/bertalign>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：sentence-transformers、faiss、sentence-splitter（無料）
- ツールの概要：文をベクトルに変換し、ベクトル同士の近さに基づいてアライメントを行う。

Vecalign

- 利用条件：Apache License 2.0
- 入手方法：<https://github.com/thompsonb/vecalign>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの概要：ベクトルベースのアライメント手法。100言語以上に対応している。

また、後述する試作コーパス作成でも採用するように、①文のベクトル化（LASER、

LaBSE、SONAR 等)、②文ベクトル間の類似度 (距離) 計算 (Faiss 等)、③類似度に基づく対訳文の抽出を、それぞれ別のツール・手法で実施する場合もある。

I-c. 機械翻訳を用いる方法

- 概要：一方の文を機械翻訳によりもう一方の言語に翻訳し、同じ言語内で類似度を計算する方法。辞書を用いる方法の特別なケースと考えることもできる。
- 利点：対訳辞書を用意する必要がない。
- 課題：機械翻訳の精度によって得られる対訳コーパスの質が左右される。また翻訳結果が正しくても、使われた単語が異なると類似度が低くなってしまう。
- ツール：

Bleualign

- 利用条件：GNU General Public License v2.0
- 入手方法：<https://github.com/rsennrich/Bleualign>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの概要：機械翻訳の自動評価手法 BLEU に基づくアライメント手法。どちらかの言語を機械翻訳しておく必要がある。

上記の他にも、文ではなく部分的な対応を抽出する方法もある。

SpanAlign

- 利用条件：NTT License (<https://github.com/nttcslab-nlp/spanalign/blob/main/LICENSE.txt>)
- 入手方法：<https://github.com/nttcslab-nlp/spanalign>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの概要：文ではなく、対応している部分を抽出することができるツール。文分割が不正確な言語などで有用な可能性がある。

5.7. 対訳辞書作成ツール

本節では対訳辞書作成ツールについて報告する。なおここでは報告するものは、対訳コーパスが与えられているという前提でのツールである。

mgiza

- 利用条件：GNU GENERAL PUBLIC LICENSE
- 入手方法：<https://github.com/moses-smt/mgiza>

- プログラミング言語：C++
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：事前の単語分割が必要。単語単位での辞書しか作成できない。

Moses

- 利用条件：LGPL
- 入手方法：<https://github.com/moses-smt/mosesdecoder>
- プログラミング言語：C++
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：Boost (無料)
- ツールの課題：インストール手順や実行手順がやや難解。Moses 自体は統計的機械翻訳を行うツールであるが、その副産物として、フレーズテーブルと呼ばれるものが生成される。これは二言語間で複数単語と複数単語との間の翻訳確率などを記録したものであるため、これを辞書として利用することが可能である。

5.8. パテントファミリーデータからの対訳コーパス作成手順

1. データの取得

上記 5.2.節で示した方法により、各言語のデータを取得する。

2. データのテキスト化

必要であれば OCR 処理や不要な箇所の削除などを行い、テキストデータに変換する。

3. 各言語の文分割

上記 5.4.節で示したツールを用いて、テキストデータを 1 文単位に文分割する。なお日本語の文分割ツールには kuzukiri もしくは GiNZA を用いる。

4. 各文のベクトル化

言語によらず共通のベクトル空間上へのベクトルに変換することができる LASER (Language-Agnostic Sentence Representations) を利用して、各言語の各文をベクトル化する。

5. 文ベクトル間の類似度 (距離) 計算

ベクトル間の類似度を高速に計算することができる Faiss を利用して各言語の全ての文間の類似度 (距離) を計算する。距離はユークリッド距離により計算される。

6. 類似度 (距離) に基づく対訳文の抽出

簡単に利用できる既存のツールが見つからなかったため、Python のスクリプトを作成した。手法は以下に示す通り。

- ・類似度は文が長いものほど信頼性が高まる傾向にあるため、日本語文の文字数が 50、40、30、20、10、0 文字以上の順に対訳文候補を選定するようにした。

- ・上記文字数以上のものについて、類似度が最も高い (距離が最も近い) 文ペアから順に、対訳文として抽出する。ただし、選定する文は出現順通りになるように制約をかけており、

文の順序が逆転してしまう場合には対訳文候補とはしない。また距離は最も近いものから上位3つまでを対訳候補とした。

・すでに対応がついている文と同じ文が対訳文候補となる場合には、前後2文以内ならば対応可とした。この際、対訳文は1対1ではなく、多対多となる。

以下に利用したツールについて紹介する。

kuzukiri

- 利用条件：MIT
- 入手方法：<https://github.com/alinear-corp/kuzukiri>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：PyO3、`unicode_normalization crate` (無料)
- ツールの課題：小数点のピリオドが文区切りとわかれてしまう。

GiNZA

- 利用条件：MIT
- 入手方法：<https://megagonlabs.github.io/ginza/>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

LASER

- 利用条件：BSD
- 入手方法：<https://github.com/facebookresearch/LASER>
- プログラミング言語：Python
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

Faiss

- 利用条件：MIT
- 入手方法：<https://github.com/facebookresearch/faiss>
- プログラミング言語：C++ & Python ラッパー
- ツールのコスト：無料
- ツールと依存関係になる他のソフトおよびそのコスト：なし
- ツールの課題：なし

<備考>

対訳文のアライメント手法として、4.~6.の手順とした理由は、Wikipedia からの多言語対訳コーパス構築²³⁴や Web データからの対訳コーパス構築²³⁵などで LASER および Faiss が利用されているなど実績があるためである。また対訳文抽出においては動的計画法が採用されることが多いが、上記の通りすぐに使えるようなツールがなく、自作するにもコストが高かったため、今回は貪欲法をベースとした方法を採用した。動的計画法を採用する場合には、例えば <https://github.com/shyyhs/CourseraParallelCorpusMining> で公開されているようなツールが参考になる。

5.9. コーパス試作、精度評価、作成可能な対訳コーパス量

5.9.1. 試作コーパスのアライメント精度評価の概要

パテントファミリーを利用してコーパスを大規模に作成する際の技術的な情報や課題について調査するため、インドネシア語、ベトナム語、タイ語の 3 言語の特許公報を対象に、日本語との対訳コーパスを試験的に作成してアライメント精度の評価を実施した。

今回の試験的なコーパス作成に使用したデータと実施した前処理の概要は以下の通りである。

インドネシア

- 使用したデータ：インドネシア/日本公報 4 件
ID2022/06891A (JP2022-538711A)、ID2022/04294A (JP2023-524548A)、ID2022/07247A (JP2023-525533A)、ID2022/08009A (JP2023-527636A)
- 対象国公報のデータ形式：透明テキスト付き PDF
- 前処理：透明テキスト付き PDF のテキストに含まれる不要部分（改行、頁番号、行番号等、表）を削除した
- インドネシア語の文分割に適用したツール：spaCy
- 日本語の文分割に適用したツール：GiNZA

ベトナム

- 使用データ：ベトナム/日本公報 6 件
VN36441A (JP2014-108812A)、VN36288A (JP2018-127746A)、VN36451A (JP2019-216575A)、VN36275A (JP2018-151331A)、VN35867A (JP2018-004113A)、VN35544A (JP2020-186347A)
- 対象国公報のデータ形式：テキストデータ

²³⁴ <https://github.com/facebookresearch/LASER/blob/main/tasks/WikiMatrix/README.md>

²³⁵ <https://ai.meta.com/blog/ccmatrix-a-billion-scale-bitext-data-set-for-training-translation-models/>

- 前処理：テキストに含まれる不要部分（表等）を削除した
- ベトナム語の文分割に適用したツール：Underthesea
- 日本語の文分割に適用したツール：GiNZA

タイ

- 使用データ：タイ/日本公報 7 件
TH147530A（JP2015-502286A）、TH147960A（JP2015-528043A）、TH152570A（JP2014-522761A）、TH141758A（JP2013-032872A）、TH150132A（JP2015-047387A）、TH170371A（JP2017-511457A）、TH1601004725A（JP2017-512014A）
- 対象国公報のデータ形式：画像 PDF
- 前処理：Adobe Acrobat を使用して画像 PDF から不要箇所（ヘッダー/フッター、行番号、ページ番号等）を取り除き、OCR ツールの FineReader を使用して画像 PDF を OCR 化し、不要部分（表等）を削除した
- タイ語の文分割に適用したツール：PyThaiNLP に含まれる sent_tokenize
- 日本語の文分割に適用したツール：kuzukiri

上記 5.8. 節「パテントファミリーデータからの対訳コーパス作成手順」で示した方法により、インドネシア語、ベトナム語、タイ語の 3 言語と対応する日本語との対訳コーパスを作成した。

なお、日本語の文分割に適用したツールとして、インドネシア語及びベトナム語の試作では GiNZA、タイ語の試作では kuzukiri を使用した。これは、作業のスケジュール上、まずは kuzukiri を利用してタイ語での試作を開始したが、その過程で kuzukiri は小数点で文分割を行ってしまうという不都合があったため、残りの二言語ではより精度が高い GiNZA を用いた。ただし、GiNZA では明かに文区切りではないところで区切ってしまうことがあるという問題もあり、一長一短であった。

5.9.2. アライメント精度評価の実施方法

5.8. 節に記載の方法により対訳コーパスを作成した際に、対応がとれていると判定された対訳文に○マークを付し、○マークを付した対訳文の中からランダムに 1 カ国あたり 200～300 文を選択して精度評価の対象とした。なお対応のついた対訳文にはスコア（距離）の情報も付与している。

○マークを付した対訳文について、対象言語を専門とするリンギストがアライメントの精度を確認し、以下の 3 段階の評価付けを行った。

表 45 アライメント精度評価の判定基準

| 評価記号 | 判定基準 |
|------|-------------------------|
| A | 一文内の 90%以上の範囲で両言語が一致する |
| B | 一文内の 21-89%の範囲で両言語が一致する |
| C | 一文内の 20%以下の範囲で両言語が一致する |

各言語に対して以下の実績を有するリングスト 1 名がアライメント精度の評価作業を行った。

表 46 アライメント精度評価を担当したリングストの経歴

| 担当言語 | 母国語 | 翻訳歴 | 経歴及び技術文書/特許文書の業務経験 |
|--------|-------|------|--|
| インドネシア | 日本語 | 14 年 | IT 関連会社及びメーカーにて提案書、仕様書や特許明細書等の翻訳業務を経験。現在はフリーランスとしてインドネシア語-日本語/英語の翻訳業務と日本語-英語のエディター業務に従事。主な専門分野は機械、IT、繊維。 |
| ベトナム | ベトナム語 | 6 年 | ベトナムの日本語学部を卒業後、翻訳会社で翻訳業務に従事。IT 通信、機械分野を中心にマニュアルや論文等の技術文章の翻訳を多数経験。特許明細書の翻訳実績も有する。 |
| タイ | 日本語 | 20 年 | 外資系企業にて勤務した後、フリーランスとしてタイ語-日本語/英語の通訳・翻訳業務に従事。化学、IT 通信、機械等分野の通訳・翻訳を経験。特許に関しては抄録、特許明細書の翻訳を経験。 |

評価作業前のフォーマット

アライメントが正しくなされていない場合、リングストは、日本語と対象言語とで一致しない範囲に赤字を付すものとし、また、対象行で一致しない部分（赤字部分）の対訳となる箇所が、その直前または直後の行に配置されている場合は、対象外の行内にある対訳部分に緑字を付すものとした。

| 1 | 日本語の 文数 | 日本語 | 自動判定 結果 | 対象言語の 文数 | 日本語文数 -対象言語 文数 | 対象言語 | 人手精原 評価(高A 中B,低C) | 日本語の 文分割が × | 対象言語 の文分割 が× | コメント | 文字化け有り | 対象 |
|----|------------|---|------------|-------------|----------------------|---|-------------------------|-------------------|--------------------|----------|--------|----|
| 78 | 1 | MERSも、ヒトにおける深刻で多くの場合に致死的な呼吸器疾患であり、MERS-CoVに指定された新型コロナウイルスによって引き起こされる。 | ○ | 1 | 1-1 | MERS juga merupakan penyakit pernapasan yang serius dan sering kali fatal pada manusia yang disebabkan oleh virus corona baru, yang disebut MERS-CoV. MERS pertama kali ditemukan di Jazirah Arab dan diyakini berasal dari unta. | B | | × | 79の対訳です。 | | 対象 |
| 78 | | MERSは、最初、アラビア半島で見つかり、ラクダに起源を持つと考えられた。 | | | | | | | | | | |
| 80 | | 2012年9月から2015年6月5日までの間に、1279人の症例と、496人の死亡、または38.7%の死亡率が報告されている。 | ○ | | | Antara September 2012 dan 5 Juni 2015, telah dilaporkan 1279 kasus MERS pada manusia dan 495 kematian, atau tingkat kematian 38,7%. | | | | | | |

図 39 対訳コーパスの評価作業用フォーマットの例

日本語と対象言語それぞれの文の数を2列に分けて記入し、最終的に統合して「日本語-対象言語」の文数の組合せを1-1、1-2、1-3、2-1、2-2等と表記した。(上記フォーマットの左から2列目と5列目を統合し、6列目に「1-1」等と記載した。)

アライメント精度の評価が低くなる要因として、日本語と対象言語のいずれかまたは両方の文分割に誤りがあることが挙げられるため、アライメント精度の評価が低くなる要因を特定できるよう、日本語と対象言語で文分割に誤りがあった場合には、該当する列にそれぞれ「×」と記入するものとした。(上記フォーマットの右から5列目、6列目参照。)。なお、文分割に誤りがある場合には、必然的にリングストの評価も悪くなるといえる(他言語も同様)。

アライメント済データにおいて文分割に誤りがあると判定した場合の主な例を以下に2点記載する。

- 一方の言語において1つの文が複数の行に分かれている

例)

「米国特許公報第6, 372, 760号に記載された安定化剤のうちのクエン酸、酒石酸又は安息香酸を含む経皮吸収製剤をそれぞれ比較例5～7とし、～それぞれ比較例8～16とした。」という一文が、アライメント済データにおいて「米国特許公報」、「第6, 372, 760号に記載された」、「安定化剤のうちのクエン酸、酒石酸又は安息香酸を含む経皮吸収製剤をそれぞれ比較例5～7とし、～それぞれ比較例8～16とした。」の3行に分割されている。

- 一方の言語における1つの文がもう一方の言語の2つ以上の文に対応している

例)

下記図において、日本語は2行目の1つのセルに文全体が記載されているが、対象言語ではその対訳が1行目の文と2行目の文とに分かれている。

| 日本語の文数 | 日本語 | 自動判定結果 | 対象言語の文数 | 対象言語 | 人手精査評価 (高A・中B・低C) | 日本語の文分割が× | 対象言語の文分割が× | コメント |
|--------|---|--------|---------|--|----------------------|-----------|------------|------------------------|
| 1 | 【0059】 また、以上では、乾燥部11におけるごみの発熱量に基づいて給じん装置40及び空気供給装置60を制御する場合について説明したが、乾燥部11におけるごみの発熱量に基づいて乾燥ストーク15、燃焼ストーク20、及び後燃焼ストーク21の全部又は一部を制御してもよい。 | ○ | 1 | Phần mô tả nêu trên mô tả trường hợp trong đó bộ cấp chất thải 40 và thiết bị cấp không khí 50 được điều khiển dựa trên năng suất tỏa nhiệt của chất thải trong máy làm khô 11. Ngoài ra, máy nạp liệu làm khô 15, máy nạp liệu đốt 20, và máy nạp liệu sau cháy 21 có thể được điều khiển toàn bộ hoặc một phần dựa trên năng suất tỏa nhiệt của chất thải trong máy làm khô 11. | C | | x | 315-316番の対象言語の対訳になっている |

図 40 対象言語の文分割に誤りがある場合の例

アライメント精度の評価を行うリングストは、特記事項がある場合に所定欄にコメントを記載するものとし、また、対象言語に文字化けが発生していた場合には所定欄に文字化けの有無を記載して報告するものとした。(図 39 の右から2列目、3列目参照。)

評価作業後のフォーマット

リングストによる評価作業が完了した後、評価対象となった文のみを1言語につき1ファイルに統合し、必要な情報を追加して集計を行った。

| 注冊 | ファイル名 | No. | 日本語の文数 | 日本語 | 自動判定結果 | 対象言語の文数 | 日本語文数-対象言語文数 | 対象言語 | 人手精査評価 (A: A中-B, 既-D) | 日本語の文数/対象言語の文数 | コメント | 文字化け有り | 対象 | 自動判定結果 | 日本語文字数 | 対象言語文字数 |
|----|-------|-----|--------|--|--------|---------|--------------|--|-----------------------|----------------|-------------|--------|----|------------|--------|---------|
| 16 | ii | 62 | 1 | 感染した動物を取り扱う場合、または感染者に接触する場合に、高い病原性の新興インフルエンザウイルスに晒される医療従事者および一般人を、より良く保護することが必要とされている。 | ○ | 1 | 1-1 | Ada kebutuhan untuk lebih melindungi petugas kesehatan dan warga sipil yang mungkin terpapar virus influenza yang baru dengan patogenitas tinggi, baik saat menangani hewan yang terinfeksi atau saat terpapar dengan orang yang terinfeksi. | A | | | | 対象 | 0187776268 | 87 | 239 |
| 17 | ii | 65 | 1 | SARSは、SARS関連コロナウイルス(SARS-CoV)と呼ばれるコロナウイルスによって引き起こされる。 | ○ | 1 | 1-1 | SARS disebabkan oleh virus corona, yang disebut coronavirus terkait SARS atau SARS-CoV. Coronavirus adalah famili virus RNA bersekolong rantai tunggal yang dibuat menjadi empat genera utama. | B | | 66の対訳となっている | | 対象 | 0187060205 | 53 | 190 |

図 41 対訳コーパスの評価結果集計用フォーマットの例

対訳コーパスを作成した際、上記 5.8. 節の「6. 類似度（距離）に基づく対訳文の抽出」に示した方法にて対象言語と日本語との間の類似度の高いものに○マークとスコアを付していたが、評価を行うリングストに予断を与えないようにするため、評価作業用のフォーマットからはスコアの列を除外していた。評価作業完了後、統合した集計用フォーマットにスコア列を追加して集計作業を行った。（図 41 の右から 3 列目参照。）

集計用フォーマットにおいて、日本語と対象言語それぞれのセルに含まれる文字数を 2 列に分けて記入した。（図 41 の右から 1 列目、2 列目参照。）

アライメント精度評価の実施結果を対象言語別に以下に記載する。

5.9.3. インドネシア語

(1) 集計結果

4 件のインドネシア公報と対応する日本公報から Excel 形式の対訳コーパスデータを作成し、アライメントツールにより両言語間の類似度が高いことを示す○マークが付したところ、各文数と割合は下記のような値になった。

表 47 対訳コーパス全文において○と判定された文の割合

| 対訳コーパス Excel の文数 | ○の数 | ○の割合 |
|------------------|-----|-------|
| 1,889 | 633 | 33.5% |

対訳コーパスデータの Excel において両言語の一方が空欄である場合も一行と見なして全体の行数を計上した。○と判定された 633 行の中から無作為にほぼ等間隔で評価対象となる 300 行を選出した。対象行の中には一文の単位に満たないものも含まれるが、便宜上、以下においては Excel の「一行」を「一文」と表記する。

300 文の対象文を A~C の判定結果別に集計した結果を以下に示す。「一文内の 90%以上の範囲で両言語が一致する」ことを示す A 判定の割合が最も多く、全体の 71.3%を占めていた。「一文内の 21-89%の範囲で両言語が一致する」ことを示す B 判定の割合は全体の 26.7%であり、「一文内の 20%以下の範囲で両言語が一致する」ことを示す C 判定の割合は全体の 2.0%であった。

表 48 人手評価の判定結果

| 人手評価の判定結果 | 該当文数 | 割合 |
|-----------|------|--------|
| A | 214 | 71.3% |
| B | 80 | 26.7% |
| C | 6 | 2.0% |
| 総計 | 300 | 100.0% |

「日本語-対象言語」の文数の組合せ別に A~C の人手評価の該当文数を集計した結果 (表及びグラフ) を以下に示す。

表 49 文数の組合せ別評価 (表)

| 文数の組合せ | A | B | C | 総計 | 割合 |
|--------|-----|----|---|-----|--------|
| 1-1 | 190 | 56 | 5 | 251 | 83.7% |
| 1-2 | 1 | 14 | | 15 | 5.0% |
| 1-3 | 1 | 6 | | 7 | 2.3% |
| 1-4 | | 1 | | 1 | 0.3% |
| 2-1 | 2 | | | 2 | 0.7% |
| 2-2 | 15 | | 1 | 16 | 5.3% |
| 2-3 | 2 | 3 | | 5 | 1.7% |
| 3-3 | 3 | | | 3 | 1.0% |
| 総計 | 214 | 80 | 6 | 300 | 100.0% |

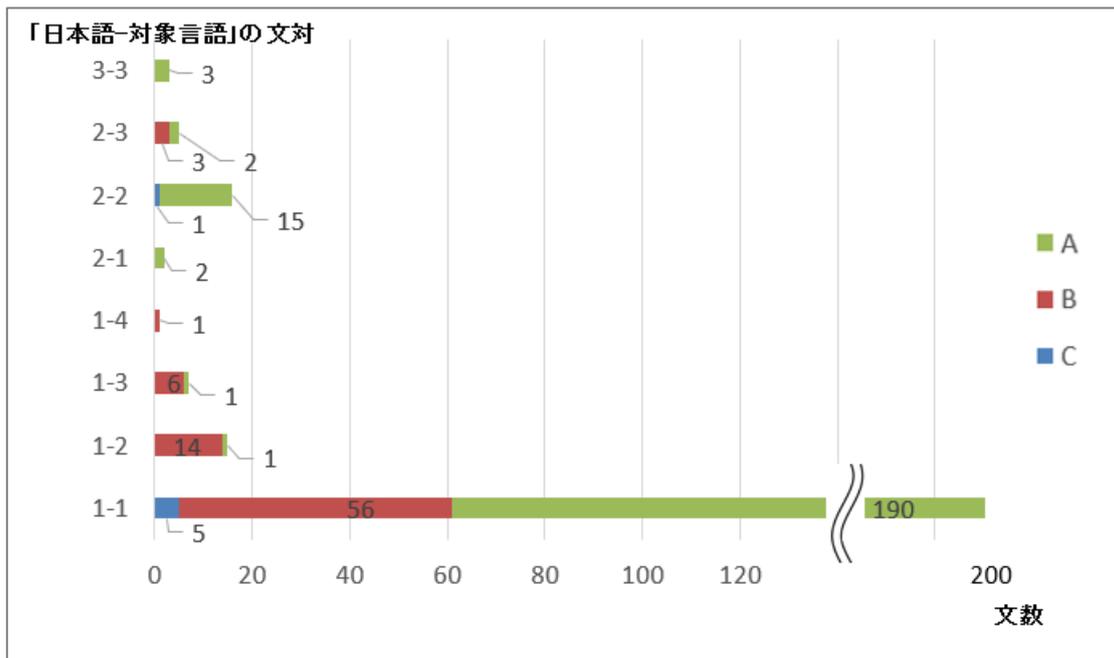


図 42 文数の組合せ別評価 (グラフ)

A 判定となった文対の大半において「日本語-対象言語」の文数の組合せは「1-1」であった。全体においても「日本語-対象言語」の文数の組合せが「1-1」であるものが最も多く、全体の83.7%を占めていた。

アライメントツールによって算出された日本語と対象言語との類似度を示すスコアと、リンギストによる人手評価との相関性について調査した。

スコアは値が小さいものほど類似度が高いことを示しており、類似度が最も低いことを示す最大値は1となる。

0-0.250 の範囲にある各文対のスコアを9段階に分け、A~Cの人手評価に該当する文数を集計した結果は以下の通りである。

表 50 スコアと人手評価結果との相関 (表)

| スコア範囲 | A | B | C | 総計 |
|-------------|-----|----|---|-----|
| 0-0.060 | 3 | | | 3 |
| 0.061-0.080 | 8 | | | 8 |
| 0.081-0.100 | 19 | 1 | | 20 |
| 0.101-0.120 | 41 | 2 | | 43 |
| 0.121-0.140 | 41 | 4 | | 45 |
| 0.141-0.160 | 40 | 19 | 1 | 60 |
| 0.161-0.180 | 37 | 23 | 1 | 61 |
| 0.181-0.200 | 16 | 14 | 4 | 34 |
| 0.201-0.250 | 9 | 17 | | 26 |
| 総計 | 214 | 80 | 6 | 300 |

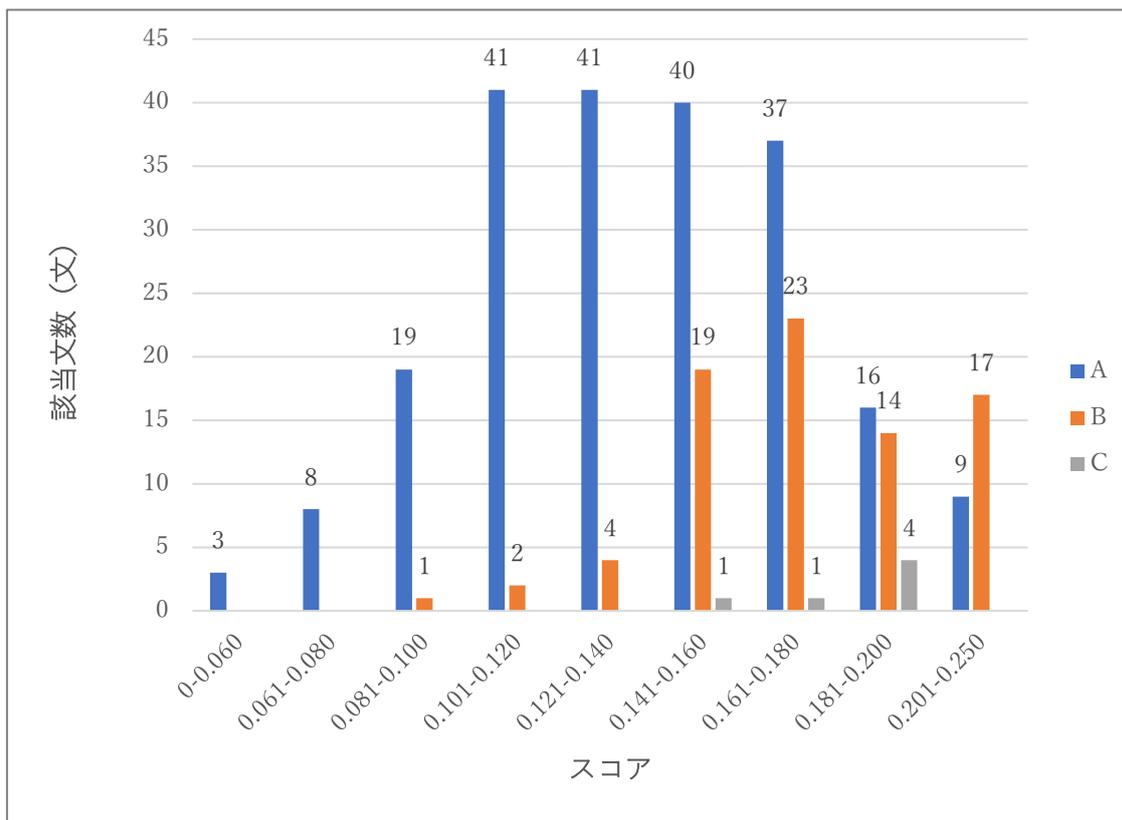


図 43 スコアと人手評価結果との相関 (グラフ)

図 43 のグラフは、横軸において、最も類似度が高いことを示すスコア 0 から最も類似度が低いことを示す 0.250 までの範囲を右方向に配置し、A~C の人手評価結果ごとに縦軸の該当文数を集計したものである。このグラフから視覚的にもわかるように、日本語の文と対象言語の文との類似度が高いスコアの 0~0.140 までの範囲に人手評価結果の A 判定が多く含まれ、類似度が低いスコア 0.161~0.200 までの範囲に C 判定が集中しており、アライメントツールから算出したスコアと人手評価の結果との間に関連性がみられることが明らかとなった。

日本語と対象言語の文分割に誤りがある文にはそれぞれの該当欄に「×」を記入したが、×の数は以下の通りとなり、文分割の誤りはインドネシア語の方がやや多いという結果となった。

日本語の文分割に誤りがある・・・43 文

インドネシア語の文分割に誤りがある・・・63 文

文分割の誤りと判定された箇所を確認したところ、日本語が長文であるときに、日本語の 1 文がインドネシア語の 2 文に対応し、そのインドネシア語の 2 文のうち的一方が前後の行に誤って配置されていることによるものが大半であった。

(2) リンギストによる総評

インドネシア公報において「C1-12」が「Cj-Cu」に変更されるといった文字化け該当箇所はほとんど検出されなかった。

全体の 71.3%が A 判定であることからわかるように、アライメントの精度は比較的に高かった。

文章が長いものほどアライメントが不正確になりやすい印象があり、短く明瞭な文章であれば自動翻訳に適しているとの報告がなされた。

(3) 機械的にコーパスを作成する方法について

アライメントツールから算出したスコアと各言語の文字数を閾値として使用し、対訳コーパス全体から利用可能な範囲を抽出することができるか否か検討した。

表 51 各人手評価におけるインドネシア語文字数とスコアとの相関

| | A | | | | | B | | | | | C | | | 総計 |
|-------------|--------|---------|---------|---------|----------|--------|---------|---------|---------|----------|--------|---------|---------|-----|
| | 40-239 | 240-439 | 440-639 | 640-839 | 840-1039 | 40-239 | 240-439 | 440-639 | 640-839 | 840-1039 | 40-239 | 240-439 | 440-639 | |
| 0-0.060 | 3 | | | | | | | | | | | | | 3 |
| 0.061-0.080 | 6 | 2 | | | | | | | | | | | | 8 |
| 0.081-0.100 | 13 | 6 | | | | | | 1 | | | | | | 20 |
| 0.101-0.120 | 29 | 7 | 5 | | | | | 1 | 1 | | | | | 43 |
| 0.121-0.140 | 22 | 16 | 2 | 1 | | 1 | 3 | | | | | | | 45 |
| 0.141-0.160 | 19 | 15 | 5 | 1 | | 5 | 9 | 2 | 2 | 1 | | 1 | | 60 |
| 0.161-0.180 | 21 | 9 | 4 | 2 | 1 | 6 | 10 | 5 | 2 | | | 1 | | 61 |
| 0.181-0.200 | 9 | 5 | 2 | | | 1 | 7 | 5 | 1 | | 3 | | 1 | 34 |
| 0.201-0.250 | 3 | 4 | 2 | | | 6 | 9 | 2 | | | | | | 26 |
| 総計 | 125 | 64 | 20 | 4 | 1 | 19 | 38 | 16 | 6 | 1 | 3 | 2 | 1 | 300 |

一文に含まれるインドネシア語の文字数を 200 文字単位でグループ化し、A~C の人手評価とスコアとの相関性を調べたところ、文字数が少ないものの中にスコアや人手評価によって高評価となる文が多いことがわかった。

仮の閾値として、スコアが 0.180 以下で、インドネシア語の文字数が 639 文字以下のものに限定して調査したところ、該当する文数は 229 文で、そのうち C 判定と評価された文は 2 文という結果になり、C 判定が発生する割合は 0.9% (2÷229×100) となった (上記表

中のオレンジの背景色で示した箇所参照)。

(4) コーパスを大規模に作成する際の課題

対訳コーパス作成の前処理として、インドネシア公報の透明テキスト付き PDF からテキストデータをコピーして Word ファイルに貼り付けたが、PDF の各行で不要な改行がなされていたため、改行処理に時間を費やす結果となった。改行処理で行う変換処理のうち、目視確認を伴わずに一括処理が可能な部分をマクロ化することにより、前処理を効率的に進めることができると思われる。

インドネシア語のテキストには数字・記号の文字化けがほとんどなく、精度は比較的高かった。大半のインドネシア語の文において、文の単位で正しく分割がなされていたが、日本語が長文である場合等に日本語の1文がインドネシア語の2文に対応し、2文のうちの一方が前後にずれてしまっていた箇所が散見された。

このようなインドネシア語の文の位置のずれを改善する方法として、文抽出アルゴリズムの変更が考えられる。具体的には、上記 5.8. 節の「6. 類似度（距離）に基づく対訳文の抽出」に記載した工程において適用する文抽出アルゴリズムを今後動的計画法に変えることにより、アライメント精度の改善が見られる可能性がある。

5.9.4. ベトナム語

(1) 集計結果

6件のベトナム公報と対応する日本公報から Excel 形式の対訳コーパスデータを作成し、アライメントツールにより両言語間の類似度が高いことを示す○マークが付したところ、各文数と割合は下記のような値になった。

表 52 対訳コーパス全文において○と判定された文の割合

| 対訳コーパス Excel の文数 | ○の数 | ○の割合 |
|------------------|-----|-------|
| 2,930 | 710 | 24.2% |

対訳コーパスデータの Excel において○と判定された 710 行の中から無作為にほぼ等間隔で評価対象となる 200 行を選出した。

200 文の対象文を A~C の判定結果別に集計した結果を以下に示す。「一文内の 90%以上の範囲で両言語が一致する」ことを示す A 判定の割合が最も多く、全体の 68.5%を占めていた。「一文内の 21-89%の範囲で両言語が一致する」ことを示す B 判定の割合は全体の 15.0%、「一文内の 20%以下の範囲で両言語が一致する」ことを示す C 判定の割合は全体の 16.5%とほぼ同程度の割合であった。

表 53 人手評価の判定結果

| 人手評価の判定結果 | 該当文数 | 割合 |
|-----------|------|--------|
| A | 137 | 68.5% |
| B | 30 | 15.0% |
| C | 33 | 16.5% |
| 総計 | 200 | 100.0% |

「日本語-対象言語」の文数の組合せ別に A~C の人手評価の該当文数を集計した結果 (表及びグラフ) を以下に示す。

表 54 文数の組合せ別評価 (表)

| 文数の組合せ | A | B | C | 総計 | 割合 |
|--------|-----|----|----|-----|--------|
| 1-1 | 129 | 23 | 22 | 174 | 87.0% |
| 1-2 | 2 | 5 | 7 | 14 | 7.0% |
| 1-3 | | 2 | 2 | 4 | 2.0% |
| 2-2 | 5 | | 1 | 6 | 3.0% |
| 2-3 | | | 1 | 1 | 0.5% |
| 3-3 | 1 | | | 1 | 0.5% |
| 総計 | 137 | 30 | 33 | 200 | 100.0% |

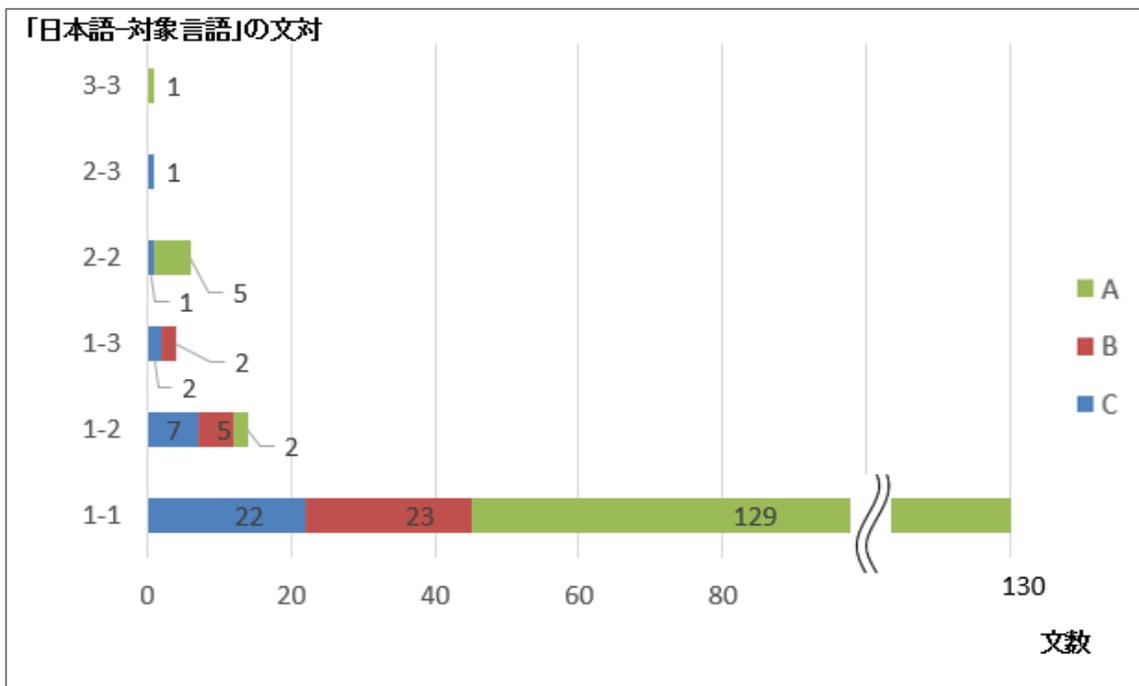


図 44 文数の組合せ別評価 (グラフ)

A 判定となった文対の大半において「日本語-対象言語」の文数の組合せは「1-1」であ

った。全体においても「日本語-対象言語」の文数の組合せが「1-1」であるものが最も多く、全体の87.0%を占めていた。

アライメントツールによって算出された日本語と対象言語との類似度を示すスコアと、リンギストによる人手評価との相関性について調査した。

スコアは値が小さいものほど類似度が高いことを示しており、類似度が最も低いことを示す最大値は1となる。

0-0.220 の範囲にある各文対のスコアを7段階に分け、A~Cの人手評価に該当する文数を集計した結果は以下の通りである。

表 55 スコアと人手評価結果との相関 (表)

| スコア範囲 | A | B | C | 総計 |
|-------------|-----|----|----|-----|
| 0-0.100 | 11 | 1 | 2 | 14 |
| 0.101-0.120 | 47 | 3 | 4 | 54 |
| 0.121-0.140 | 49 | 12 | 10 | 71 |
| 0.141-0.160 | 24 | 10 | 11 | 45 |
| 0.161-0.180 | 6 | 4 | 3 | 13 |
| 0.181-0.200 | | | 2 | 2 |
| 0.201-0.220 | | | 1 | 1 |
| 総計 | 137 | 30 | 33 | 200 |

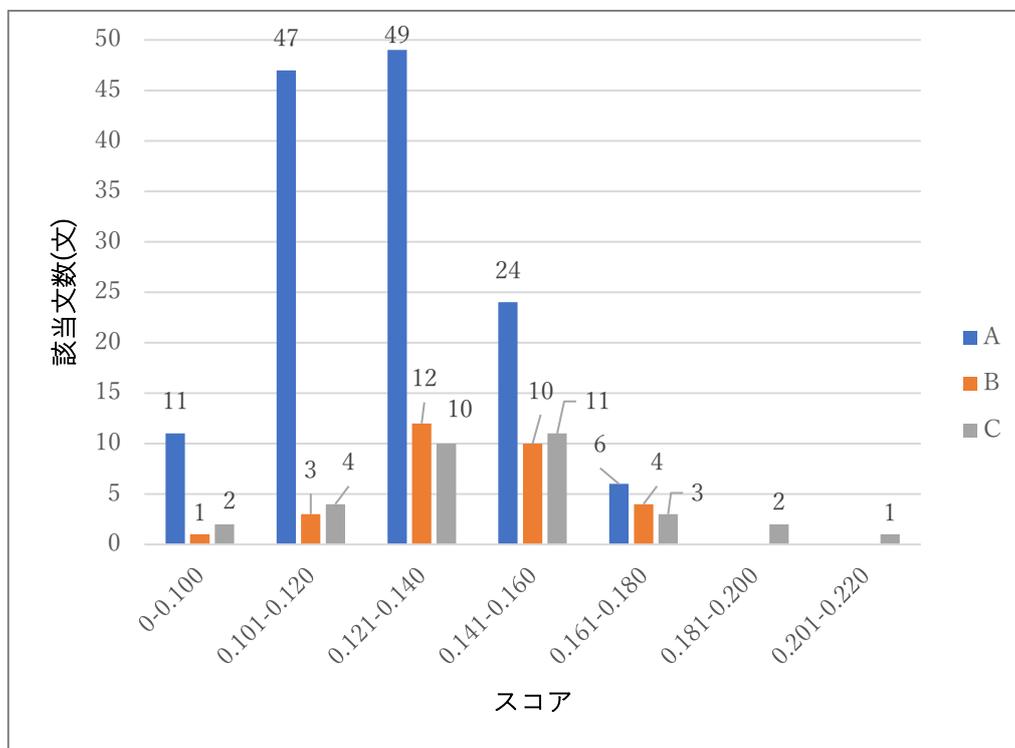


図 45 スコアと人手評価結果との相関 (グラフ)

図 45 のグラフは、横軸において、最も類似度が高いことを示すスコア 0 から最も類似度が低いことを示す 0.220 までの範囲を右方向に配置し、A～C の人手評価結果ごとに縦軸の該当文数を集計したものである。このグラフから視覚的にもわかるように、日本語の文と対象言語の文との類似度が高いスコアの 0～0.140 までの範囲に人手評価結果の A 判定が多く含まれ、類似度が低いスコア 0.141～0.220 までの範囲に C 判定が集中しており、アライメントツールから算出したスコアと人手評価の結果との間に関連性がみられることが明らかとなった。

日本語と対象言語の文分割に誤りがある行にはそれぞれの該当欄に「×」を記入したが、×の数は以下の通りとなり、文分割の誤りはベトナム語の方がやや多いという結果となった。

日本語の文分割に誤りがある・・・26 行

ベトナム語の文分割に誤りがある・・・32 行

文分割の誤りと判定された箇所を確認したところ、以下が主な要因となっていた。

- ・図番号中の「Fig. x」のピリオドが文末を示すピリオドとして扱われている。
- ・日本語の 1 文がベトナム語の 2 文に対応し、2 文のうち的一方が前後に配置されている。
- ・日本語、ベトナム語共に Excel の 1 セルの中に複数の文が配置されている。

(2) リンギストによる総評

ベトナム公報のテキストにおいて「trước」(誤) → 「trước」(正)、「Két」(誤) → 「Ket」(正) 等の誤記がいくつか検出されたが、文字化けと思われるものはほとんどなかった。

全体の 68.5%が A 判定であることからわかるように、アライメントの精度は比較的に高かった。

日本語の文末には句点（「。」）があり、ベトナム語の文末にはピリオド（「.」）があるため、文末記号が明確であるにもかかわらず、一行に 2 文以上が入り込んでいる箇所が複数あり、改善の余地があるように思われた。

(3) 機械的にコーパスを作成する方法について

アライメントツールから算出したスコアと各言語の文字数を閾値として使用し、対訳コーパス全体から利用可能な範囲を抽出することができるか否か検討した。

表 56 各人手評価におけるベトナム語文字数とスコアとの相関

| | A | | | | B | | | C | | | 総計 |
|-------------|------------|-------------|-------------|-------------|------------|-------------|-------------|------------|-------------|-------------|-----|
| | 50- 249 | 250- 449 | 450- 649 | 650- 849 | 50- 249 | 250- 449 | 450- 649 | 50- 249 | 250- 449 | 450- 649 | |
| 0-0.090 | 1 | 2 | | | | | | | 1 | | 4 |
| 0.091-0.110 | 16 | 11 | | | 1 | 1 | | | 2 | | 31 |
| 0.111-0.130 | 35 | 18 | 3 | | 2 | 4 | 1 | 2 | 4 | | 69 |
| 0.131-0.150 | 22 | 9 | | 1 | 4 | 4 | 2 | 6 | 4 | 1 | 53 |
| 0.151-0.170 | 6 | 9 | 1 | | 4 | 3 | 1 | 4 | 4 | | 32 |
| 0.171-0.190 | 3 | | | | 2 | 1 | | 2 | 2 | | 10 |
| 0.191-0.210 | | | | | | | | 1 | | | 1 |
| 総計 | 83 | 49 | 4 | 1 | 13 | 13 | 4 | 15 | 17 | 1 | 200 |

一文に含まれるベトナム語の文字数を 200 文字単位でグループ化し、A～C の人手評価とスコアとの相関性を調べたところ、文字数が少ないものの中にスコアや人手評価によって高評価となる文が多いことがわかった。

仮の閾値として、スコアが 0.170 以下で、ベトナム語の文字数が 249 文字以下のものに限って調査したところ、該当する文数は 103 文で、そのうち C 判定と評価された文は 12 文という結果になり、C 判定が発生する割合は 11.7% (12÷103×100) となった (上記表中のオレンジの背景色で示した箇所参照)。

(4) コーパスを大規模に作成する際の課題

今回のコーパス試作において、ベトナム語のテキストはベトナム国家知的財産庁 (IP VIET NAM) のウェブサイトから直接コピーしたものを使用し、頁番号や行番号を削除する必要もなく、前処理には比較的工数がかからなかった。しかし、「5.2.6. ベトナム語」の節に記載したように、ウェブサイトから全文のテキストが取得できる公報とできない公報があるため、多数の公報を収集する際には画像 PDF の OCR 化等の処理も検討する必要がある。

上記のウェブサイトから取得したベトナム語のテキストには数字・記号の文字化けがほとんどなく、精度は比較的高かった。インドネシア語と同様に、ベトナム語においても日本語が長文である場合等に日本語の 1 文がベトナム語の 2 文に対応し、2 文のうち的一方が前後にずれてしまっていた箇所が見られた。

このようなベトナム語の文の位置のずれを改善する方法として、インドネシア語と同様に、文抽出アルゴリズムを動的計画法に変えることが挙げられる。

5.9.5. タイ語

(1) 集計結果

7 件のタイ公報と対応する日本公報から Excel 形式の対訳コーパスデータを作成し、アライメントツールにより両言語間の類似度が高いことを示す○マークが付したところ、各文数と割合は下記のような値になった。

表 57 対訳コーパス全文において○と判定された文の割合

| 対訳コーパス Excel の文数 | ○の数 | ○の割合 |
|------------------|-----|-------|
| 3,410 | 404 | 11.8% |

対訳コーパスデータの Excel において○と判定された 404 行の中から無作為にほぼ等間隔で評価対象となる 300 行を選出した。

300 文の対象文を A～C の判定結果別に集計した結果を以下に示す。「一文内の 90%以上の範囲で両言語が一致する」ことを示す A 判定の割合は全体の 16.7%であった。「一文内の 21-89%の範囲で両言語が一致する」ことを示す B 判定の割合は全体の 71.7%であり、大半が B 判定に該当する結果となった。「一文内の 20%以下の範囲で両言語が一致する」ことを示す C 判定の割合は最も少なく、全体の 11.7%であった。

表 58 人手評価の判定結果

| 人手評価の判定結果 | 該当文数 | 割合 |
|-----------|------|--------|
| A | 50 | 16.7% |
| B | 215 | 71.7% |
| C | 35 | 11.7% |
| 総計 | 300 | 100.0% |

「日本語-対象言語」の文数の組合せ別に A～C の人手評価の該当文数を集計した結果（表及びグラフ）を以下に示す。

表 59 文数の組合せ別評価（表）

| 文数の組合せ | A | B | C | 総計 | 割合 |
|--------|----|-----|----|-----|--------|
| 1-1 | 50 | 124 | 25 | 199 | 66.3% |
| 1-2 | | 56 | 5 | 61 | 20.3% |
| 1-3 | | 19 | 2 | 21 | 7.0% |
| 1-4 | | 5 | 2 | 7 | 2.3% |
| 1-5 | | 2 | | 2 | 0.7% |
| 2-1 | | | 1 | 1 | 0.3% |
| 2-2 | | 6 | | 6 | 2.0% |
| 2-3 | | 2 | | 2 | 0.7% |
| 3-3 | | 1 | | 1 | 0.3% |
| 総計 | 50 | 215 | 35 | 300 | 100.0% |

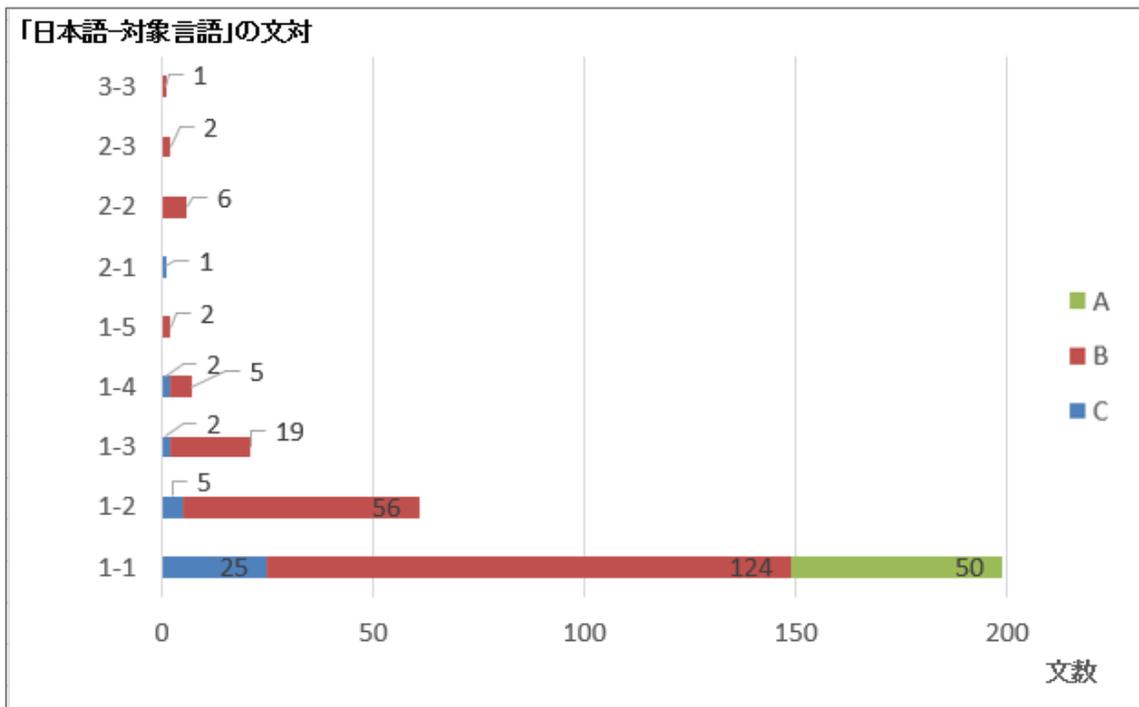


図 46 文数の組合せ別評価（グラフ）

A 判定となった文対のすべてにおいて「日本語-対象言語」の文数の組合せは「1-1」であった。全体においても「日本語-対象言語」の文数の組合せが「1-1」であるものが最も多く、全体の 66.3%を占めていた。

アライメントツールによって算出された日本語と対象言語との類似度を示すスコアと、リングストによる人手評価との相関性について調査した。

スコアは値が小さいものほど類似度が高いことを示しており、類似度が最も低いことを示す最大値は1となる。

0-0.250 の範囲にある各文対のスコアを 8 段階に分け、A～C の人手評価に該当する文数を集計した結果は以下の通りである。

表 60 スコアと人手評価結果との相関 (表)

| スコア範囲 | A | B | C | 総計 |
|-------------|----|-----|----|-----|
| 0-0.100 | 6 | 6 | | 12 |
| 0.101-0.120 | 18 | 26 | | 44 |
| 0.121-0.140 | 13 | 64 | 9 | 86 |
| 0.141-0.160 | 11 | 74 | 13 | 98 |
| 0.161-0.180 | 1 | 31 | 7 | 39 |
| 0.181-0.200 | 1 | 13 | 4 | 18 |
| 0.201-0.220 | | | 2 | 2 |
| 0.221-0.250 | | 1 | | 1 |
| 総計 | 50 | 215 | 35 | 300 |

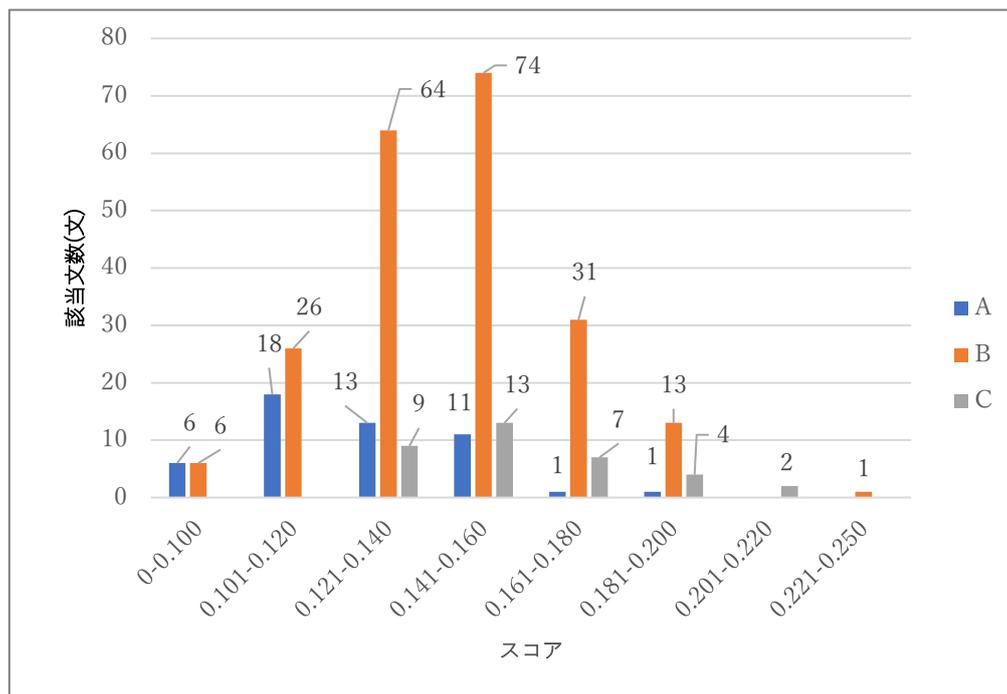


図 47 スコアと人手評価結果との相関 (グラフ)

図 47 のグラフは、横軸において、最も類似度が高いことを示すスコア 0 から最も類似度が低いことを示す 0.250 までの範囲を右方向に配置し、A～C の人手評価結果ごとに縦軸の該当文数を集計したものである。このグラフから視覚的にもわかるように、日本語の文と対象言語の文との類似度が高いスコアの 0～0.140 までの範囲に人手評価結果の A 判定が多

く含まれ、類似度が低いスコア 0.141~0.220 までの範囲に C 判定が集中しており、アライメントツールから算出したスコアと人手評価の結果との間に関連性がみられることが明らかとなった。

日本語と対象言語の文分割に誤りがある行にはそれぞれの該当欄に「×」を記入したが、×の数は以下の通りとなり、文分割の誤りは大半がタイ語において発生していることがわかった。

日本語の文分割に誤りがある・・・42 行

タイ語の文分割に誤りがある・・・232 行

| 1 | 日本語の 文数 | 日本語 | 自動判定 結果 | 対象言語 の文数 | 対象言語 | 人手評価評価 (高A中B 低C) | 日本語の文 分割数× | 対象言語の 文分割数× | コメント | 文字化け有り | 対象 |
|-----|------------|--|------------|-------------|--|------------------------|---------------|----------------|----------------------------------|--|----|
| 100 | 2 | また、前記樹脂ボードは未成形部分の範囲に重ねられていて毒度者から見えないので、外観品質を向上させることができます。 そして、樹脂ボードを挿入するボウナ部も前記成形部分に設ける構造に比べて、シートパンを製作する際の作業工程を少なくすることができ、作業性を向上させることができます。 | ○ | 2 | รับไปจากนี้ เนื่องจากแบบรูปที่ดูแล้วดูเหมือนต้นตำรับของ ส่วนของวัสดุชนิด ถูกมองไม่เห็นโดยผู้รับ ส่วนนี้ถูกผูกสภาพในรูปที่กดขึ้นได้ยากจึงสามารถกด ทำได้มีขึ้นได้ จำนวนน้อยจนยาก ในการทำงานสำหรับภายในหรือพื้นที่กลางแจ้งสามารถมองได้ก็เหมือนไม่มีความสามารถในภาพทำงาน | B | × | × | C10日本語赤マーカー部に対応するタイ語は、F10緑マーカー部。 | 【ので(เนื่องจาก)文字化け正しくは、เนื่องจาก)】 【の記載の緑マーカー部にも文字化け有り】 | 対象 |
| 103 | | (請求項 1) | | | | | | | | | |
| 104 | | [0007] | | | โปรดรับแบบกับที่ส่งสำหรับที่รับส่งและเปิดขึ้นและเปิดจะ ดูได้เข้าไปในสวน ส่วนกลางอยู่ถัดไปในส่วนของวัสดุต้นตำ | | | | | | |
| 105 | | | | | | | | | | | |

図 48 タイ語の対訳コーパスにおける文分割の誤りの例

上記の図 48 の例では、2 行目に赤字で示した日本語の範囲の対訳が 5 行目に緑字で示したタイ語の範囲と対応しており、日本語の 1 文中の一フレーズのみが対象行から数行離れた行に配置されていた。

インドネシア語、ベトナム語に比べ、タイ語の対訳コーパスでは文分割の誤りがかなり多く発生し、上記の図 48 に示したように、パターン化が困難なほど複雑な形での誤りが多数散見された。

タイ語の対訳コーパスにおいて文分割の誤りが多くなる要因として、以下のような言語特性が挙げられる。

- ・タイ語にはピリオドもカンマもなく、公文書以外の文書では文間にスペースない場合もある
- ・英数記号とタイ語の間や、「A, B, C and D」の ABCD 間に慣用的にスペースが入るため、スペースを文末記号として扱うことができない

文法規則に従って機械的にタイ語の文を分割しようとしても精度に限界があると思われるため、正確にタイ語の文を分割する際には、タイ語を解するリングストが文の間に読点等を入れる等といった人手による前処理が必要となる。

(2) リングストによる総評

タイ語の文字化けについて多くの指摘がなされ、300 文のうちの 206 文において文字化けが含まれているとコメントされていた。特に、形の似ているタイ語子音(例えば ด と ต、บ と ป、ล และ ส 等)への文字化けが多数あり、文字化けによって子音が入れ替わることで別の意味の単語になり、文意が変わるか、または存在しない単語になる等して意味が通らな

い表記になっていたとの報告がなされた。

また、タイ語にはピリオドもカンマもなく、文間にスペースあるものもないものもあるため、特定の表記から機械的に文の単位を特定することができない点も大規模にコーパスを作成する上での大きな懸念点となり得るとの指摘もあった。

(3) 機械的にコーパスを作成する方法について

アライメントツールから算出したスコアと各言語の文字数を閾値として使用し、対訳コーパス全体から利用可能な範囲を抽出することができるか否か検討した。

表 61 各人手評価におけるタイ語文字数とスコアとの相関

| | A | | | B | | | | | C | | | | 総計 |
|-------------|-------|---------|---------|-------|---------|---------|---------|---------|-------|---------|---------|---------|-----|
| | 0-199 | 200-399 | 400-599 | 0-199 | 200-399 | 400-599 | 600-799 | 800-999 | 0-199 | 200-399 | 400-599 | 600-799 | |
| 0-0.090 | 1 | 2 | | 1 | | | | | | | | | 4 |
| 0.091-0.110 | 6 | 6 | | 7 | 1 | | | | | | | | 20 |
| 0.111-0.130 | 13 | 4 | | 31 | 18 | 5 | | | 2 | | 1 | | 74 |
| 0.131-0.150 | 11 | 1 | | 34 | 28 | 7 | 2 | 1 | 7 | 1 | 1 | 1 | 94 |
| 0.151-0.170 | 3 | 1 | 1 | 26 | 17 | 7 | | 1 | 11 | 3 | | | 70 |
| 0.171-0.190 | 1 | | | 9 | 12 | 3 | | | 2 | 3 | | | 30 |
| 0.191-0.210 | | | | 3 | 1 | | | | 1 | | | | 5 |
| 0.21-0.250 | | | | | 1 | | | | 1 | 1 | | | 3 |
| 総計 | 35 | 14 | 1 | 111 | 78 | 22 | 2 | 2 | 24 | 8 | 2 | 1 | 300 |

一文に含まれるタイ語の文字数を 200 文字単位でグループ化し、A~C の人手評価とスコアとの相関性を調べたところ、文字数が少ないものの中にスコアや人手評価によって高評価となる文が多いことがわかった。

仮の閾値として、スコアが 0.150 以下で、タイ語の文字数が 399 文字以下のものに限って調査したところ、該当する文数は 174 文で、そのうち C 判定と評価された文は 10 文という結果になり、C 判定が発生する割合は 5.7% (10÷174×100) となった（上記表中のオレンジの背景色で示した箇所参照）。

(4) コーパスを大規模に作成する際の課題

対訳コーパス作成の前処理として、OCR ツールの FineReader を適用してタイ公報の画像 PDF を OCR した。OCR データ内にヘッダー/フッター、頁番号、及び行番号等の不要な部分が入り込まないように、事前に Adobe Acrobat を使用して PDF 公報の余白のトリミングを行う等、前処理に時間を費やす結果となった。大量に処理する際には、PDF のトリミングと OCR 処理において PDF ツールの一括処理機能を活用する等、効率を上げる方策があり得る

ように思われる。

タイ語のテキストにおいて文字化けが非常に多いことがアライメントの精度を著しく損ねているとリンギストから報告が上がっていた。OCR は使用するツールによって精度が異なるので、タイ語の OCR 処理に適したツールを模索することで改善を図ることもできるが、この方法で文字化けの問題を解消することは困難であると言わざるを得ない。

正確なテキストが入手できないため、文分割ツールやアライメントツールの調整や見直しによってアライメント精度の向上を図ることも困難であるが、タイ語における末尾における明示的な記号（ピリオドや句点）のないという言語的な特性によって、文分割の精度も低い状況である。なお、この点については、上述の SpanAlign のように、文ではなく、対応している部分を抽出することができるアライメントツールが有用な可能性もある。

5.10. その他の関連技術

5.10.1. 多言語で事前訓練された大規模なモデルを用いた多言語処理について

5.8.節では多言語で事前訓練された大規模モデルとして LASER を利用した。LASER を含め、その他の事前学習モデルを以下に示す。

LASER2/LASER3

- 対応言語：LASER2 は約 100 言語、LASER3 は 147 言語、合わせて約 250 言語
- 入手方法：<https://github.com/facebookresearch/LASER/tree/main/nllb>
- ツールの特徴など：対応言語数が多い。
- 論文：<https://aclanthology.org/2022.findings-emnlp.154/>

MBart と MBart-50

- 対応言語：MBart は 25 言語、MBart-50 は 50 言語
- 入手方法：https://huggingface.co/docs/transformers/model_doc/mbart
- ツールの特徴など：英語の事前学習モデルである BART²³⁶を多言語に拡張したもの。
- 論文：MBart <https://aclanthology.org/2020.tacl-1.47>
MBart-50 <https://aclanthology.org/2021.findings-acl.304>

LaBSE

- 対応言語：109 言語
- 入手方法：<https://huggingface.co/sentence-transformers/LaBSE>
- ツールの特徴など：BERT²³⁷を多言語に拡張したもの。LASER よりも精度が良いとされ

²³⁶

<https://github.com/facebookresearch/fairseq/tree/33cfe372812f42eb6b1fb5dcc07f3f7f810c5ea/examples/bart>

²³⁷ <https://arxiv.org/abs/1810.04805>

ている。

- 論文 : <https://aclanthology.org/2022.acl-long.62/>

SONAR

- 対応言語 : 200 言語

- 入手方法 : <https://huggingface.co/facebook/SONAR>

- ツールの特徴など : 音声と言語の両方を用いて訓練されている。LASER や LaBSE よりも高精度とされている。音声からテキストへの翻訳なども行える。

- 論文 : <https://arxiv.org/abs/2308.11466>

5.10.2. 対訳コーパスのデータ拡張

Data Augmentation for Low-Resource Neural Machine Translation

- 論文 : <https://aclanthology.org/P17-2090/>

- 概要 : 文内の低頻度の語を、高頻度で意味の近い別の語に置き換えることでデータ拡張を行う手法。

RAML: Reward Augmented Maximum Likelihood for Neural Structured Prediction

- 論文 : <https://proceedings.neurips.cc/paper/2016/hash/2f885d0fbe2e131bfc9d98363e55d1d4-Abstract.html>

- 概要 : 目的言語側の文を編集し、編集距離を負の報酬として学習に用いる手法。デコーダーの汎化性能の向上が見込める。

SwitchOut: an Efficient Data Augmentation Algorithm for Neural Machine Translation

- 論文 : <https://aclanthology.org/D18-1100/>

- 概要 : 上記の RAML と似ているが、現言語側の文も置き換える点異なる。

WordDropout

- 論文 : <https://aclanthology.org/W16-2323/>

- 概要 : 原文もしくは目的言語文の任意の単語のベクトルを 0 にしてしまうことで、エンコーダーやデコーダーの汎化性能を向上させる手法。

Exploiting Source-side Monolingual Data in Neural Machine Translation

- 論文 : <https://aclanthology.org/D16-1160/>

- 概要 : 原言語の大規模単言語コーパスを用いて NMT の性能向上をはかる手法。小規模の対訳コーパスを用いて訓練した NMT により大規模単言語コーパスを翻訳することで対訳コーパスを増やす。さらに原言語の単語を並べ替えたものを元に戻すタスクとのマルチタスク学習を行うことで精度が向上することを示している。

Improving Neural Machine Translation Models with Monolingual Data

- 論文 : <https://aclanthology.org/P16-1009/>

- 概要 : 逆翻訳によるデータ拡張を行う手法。目的言語文の単言語コーパスを原言語に翻訳することで対訳コーパスの拡充を行う。目的言語文が人手により生成された綺麗なデータであるため、機械翻訳の出力が崩れるようなことはあまり起こらないという利点がある。

Data Augmentation for Neural Machine Translation using Generative Language Model

- 論文 : <https://arxiv.org/abs/2307.16833>

- 概要 : 大規模言語モデルを使って単言語コーパスからのデータ拡張を行う手法。

Generalized Data Augmentation for Low-Resource Translation

- 論文 : <https://aclanthology.org/P19-1579/>

- 概要 : 例えば英語と低資源言語との間の機械翻訳においてデータ拡張を行う際に、低資源言語と言語的に近い高資源言語を中間言語として利用することで、より良いデータ拡張を行う手法。

6. 調査結果の総合分析

本調査では、①機械翻訳エンジンの種類・内容、②日英中韓以外の言語をネイティブとする特許庁の状況、③調査対象言語を使用する国々に関する経済的状況、④調査対象言語に関するコーパス作成の技術的情報の4つの項目に対する調査を実施した。

調査項目1では、汎用機械翻訳サービス・エンジンと、特許情報に特化した機械翻訳サービス・エンジンを調査した。汎用機械翻訳サービス・エンジンは、既にあらゆる場面で活用されるようになっているが、各社が、対象言語、カスタマイズ機能、API提供等において差異のあるものを展開していることが示された。用途が機械翻訳に限定されない大規模言語モデルを用いた機械翻訳技術に関する注目論文も出ているが、同モデルに必要な計算リソースは膨大であるところ、翻訳精度の確保、業務での利便性等の運用面を踏まえると、特許や法務等、それぞれのドメインに特化した機械翻訳サービス・エンジンの需要は高まるものと推察される。

特許特化型の機械翻訳サービス・エンジンは、公的には一部の知財庁が提供しているが、そうした知財庁は、EPOのように域内の構成国から言語の多様性を配慮する必要な機関、日本・韓国のように言語上の翻訳ニーズ（日本語－英語、韓国語－英語等）が高い国の機関である（他方、米国、英国等の知財庁では、機械翻訳機能の提供は見られなかった）。

特許特化型機械翻訳の商用サービスについては、特許関連データを学習させた機械翻訳エンジンをベースとしながら、サービス形態に応じてカスタマイズ、又は機能追加を行ったエンジンを利用している。汎用機械翻訳エンジンに比較すると、特許特化型の機械翻訳エンジンの対応言語は現状限られており（主に日英中韓独仏等）、実際、今回の調査対象の10言語のうち、日本語との間で特許特化型のエンジンが提供されるのは、スペイン語とポルトガル語のみであった。

本調査で対象としているASEAN、中南米、中東地域の言語は、現状の特許特化型機械翻訳サービスではほとんど対応しておらず、汎用機械翻訳エンジンで対応せざるを得ない状況である。特許情報はその文章や用語が一般的な文章と異なる部分も多いため、特許情報の機械翻訳サービスの提供事業者からも、汎用機械翻訳エンジンを用いた場合の翻訳精度への信頼は必ずしも得られていない。しかしながら、整備された特許情報データが存在していることは、特許情報に特化した翻訳エンジンの開発に有利であり、対訳データの作成手法に関して機械翻訳の活用や文のベクトル化ツールを活用するなど、複数の方法で対応できること、また、対訳辞書がない場合でも少量の対訳データを基に翻訳エンジンの精度を高める手法も検討可能で、ASEAN言語等の言語資源が比較的少ない言語であっても特許特化型機械翻訳を実現できる可能性が高まってきていることが窺える。そして、これらの技術は、自然言語処理の技術の進展とともに日々発展が見られることから、最新の技術やツールを把握して適用を検討していくことも重要といえる。

また、ニーズに関するヒアリング調査でも、これら言語の特許情報に対する顧客からの

翻訳ニーズは、潜在的には存在しているといえ、顧客ニーズが顕在化して高まりを見せれば、民間レベルでのサービス提供・開発機運が高まる可能性もあるといえる。

調査項目 2 では、日英中韓以外の言語をネイティブとする海外の知財庁の審査官を対象にアンケート調査を実施した。知財庁の審査官の多くは、日本の審査経過情報を英語で利用する場合、EPO が提供している Espacenet、次いで PATENTSCOPE や J-PlatPat を利用している。また、汎用的な機械翻訳サービスとしては、Google の翻訳サービスを利用している審査官が多い。調査対象の知財庁の半数以上で、審査官の任用にあたっては、一定の英語力の保持を示す指標があるとされ、欧米の公報・審査経過情報を英語で確認することに対しては、多くの審査官が支障を抱いていないという結果であったが、一方で、原文が英語以外の言語（例えば、日中韓の言語）で英訳した公報・審査結果を確認することには支障の声が一定数あがった。以上の結果より、各知財庁の審査官にとっても既存の翻訳エンジンでは必ずしも十分とはいえないことが推察される。

調査項目 3 では、調査対象国の経済状況等を把握した。調査対象の 12 カ国の経済状況（GDP の推移）は、2008 年以降、COVID-19 パンデミックを経験しつつ、どの国も経済が拡大基調にある。日本との関係では、ASEAN の一部の国で、二国間の包括的戦略的パートナーシップに格上げが図られる等、経済社会の交流機運は高まっている。

本調査では、五庁で主に使用される言語（日英中韓）以外の言語を対象に、機械翻訳²³⁸を提供することを想定すると、これら件数が高い対象国については、機械翻訳サービスの潜在的なユーザーが多い国と考えられる。直近 10 年の平均特許出願件数（日本→対象国）で上位国は、インド、タイ、インドネシア、ブラジル、ベトナムであった。他方、直近 10 年の平均特許出願件数（対象国→日本）で上位国は、イタリア、スペイン、インド、タイであった。直近 10 年の平均 PPH 申請件数（日本→対象国）で上位国は、タイ、インドネシア、マレーシア、メキシコであった。

また、これらの国に進出する企業に対してヒアリング調査を行った結果、FTO 調査（侵害予防調査）や先行技術調査にて、上述するような国を対象とする場合には、十分なデータベースが整っていないことや、既存の翻訳エンジンでの翻訳精度が十分でない点が課題として挙げられた。

調査項目 4 では 10 言語を対象に、特許情報に係るコーパス作成の技術的情報として、以下の項目について調査を行い、具体的な候補手順や課題を整理した。

- パテントファミリーデータの取得元およびデータの状態
- OCR ツール
- 文分割ツール

²³⁸ 提供が期待される機械翻訳には、五庁以外の知財庁の特許情報の日本語への機械翻訳と、日本の審査情報等の現地語への機械翻訳が期待される。

- 形態素解析（単語分割）ツール
- アライメントツール
- 対訳辞書作成ツール
- パテントファミリーデータからの対訳コーパス作成手順

対象 10 言語のうち、一定数の公報テキストデータの取得が可能であり、また、これまでの調査によって明らかにした、既存の特許特化型エンジン、経済的・技術的情報、及び、有識者の見解等も合わせて考慮し、日本企業にとって機械翻訳エンジンを活用した特許情報の検索・調査ニーズが比較的高いと思われる言語の中から、インドネシア語、ベトナム語、タイ語の 3 言語を選んでコーパスの試作と精度評価を実施し、大規模な対訳コーパス作成の際に課題となり得る点について調査した。

上記対応の結果、タイ語については、取得可能な公報データが画像データであって OCR によって得られるテキストデータの精度が悪いとともに、タイ語の言語的な特性から文分割の精度も悪いとの課題があり、現在入手可能な公報データ・ツールでは、自動的な対訳コーパス作成の困難性は高いことがわかった。

一方、インドネシア語とベトナム語は、テキストデータを含む公報データが入手可能であって、一定の精度で対訳コーパスを自動的に作成できることがわかった。ただし、ベトナム語については、現時点においては、明細書を含む全文のテキストデータが取得可能な期間が限られており、十分なテキストデータが存在するとはいえないとの課題があるといえる。一方、インドネシア語は、一定の範囲の公報で、明細書を含む全文テキストデータが入手可能であるところ、同国は日本企業の進出が活発な地域の 1 つであって、現時点では日本語との特許特化型エンジンも存在しないことから、特許情報に特化した対訳コーパスを整備する有力な候補といえる。

今回、大規模に対訳コーパスを作成する際の具体的な手段および手順についても多角的に検討を行い、適用する文抽出アルゴリズムの変更等、本調査においていくつかの改善点を見出したので、今後、上記の言語の対訳コーパスを自動的に作成する場合には、これらの改善点も活かすことができるといえる。