

平成 30 年度特許庁委託事業

**平成 30 年度  
審決の日英機械翻訳に関する調査  
調査報告書**

**平成 31 年 3 月 20 日**

**株式会社サン・フレア**

## 目次

1. 調査目的	3
2. 調査の概要	4
2.1 実施体制	4
2.2 調査スケジュール	5
2.3 調査方法	5
3. 審決の機械翻訳の自動評価	14
3.1 審決の機械翻訳の自動評価の結果及び比較	14
4. 審決の機械翻訳の人手評価	22
4.1 審決の機械翻訳の人手評価の結果及び比較	23
4.2 人手評価結果の分析コメント	26
5. 自動評価を活用した審決の機械翻訳の分析	39
5.1 人手評価と自動評価との関係	39
5.2 審決文 100 件を対象とした分析	40
6. 辞書データの作成・分析	45
6.1 登録語の編集における工夫	45
6.2 辞書データ作成についての提言	46
7. 審決の日英機械翻訳文提供に向けての課題	47
7.1 審決の日英機械翻訳文の現状	47
7.2 審決の日英機械翻訳文提供に向けての提言	56
8. 付録	58
8.1 見解書コーパス	58
8.2 審決コーパス	59

## 1. 調査目的

近年、海外企業等による日本への投資を促進するための基盤整備、あるいは日本企業が海外に進出する際の進出先への情報提供等のため我が国の法制度の運用に関する情報の英語での発信が強く求められている。<sup>12</sup>この要請を受け、裁判所では判決の英訳を公表している他、特許庁においても一部の審決について、人手翻訳した英文を外部に提供している。

審決の人手翻訳については、費用等の観点から、機械翻訳技術が注目されており、特許庁では、平成 28 年 1 月から独立行政法人工業所有権情報・研修館の特許情報プラットフォーム（J-PlatPat）を通じて、審決の機械翻訳文を一般に提供している。また平成 31 年 5 月（新元号不明）には、機械翻訳の精度を向上させた新たな日英機械翻訳システム（以下、日英機械翻訳新システムという。）の導入を目指している。

近年は、統計翻訳やニューラル翻訳といった、従来のルールベース翻訳とは異なる翻訳方式が、その翻訳精度の高さから機械翻訳の技術分野で注目を集めている。しかしながら、日英機械翻訳新システムにおいてこうした新しい翻訳方式が導入された場合、審決を機械翻訳したとき、従来起こらなかった新たな問題が発生する可能性がある。

そこで本調査では、日英機械翻訳新システムに導入されうる翻訳エンジンと類似の翻訳エンジンを利用して審決の日英機械翻訳を行うための課題を明らかにし、J-PlatPat 等の日英機械翻訳サービスにおける審決の高精度な日英機械翻訳文の提供に寄与させることを目的とする。

---

<sup>1</sup> 「国際戦略に関する提言」 自民党政務調査会 司法制度調査会 民事・国際小委員会（平成 26 年 5 月 30 日）

<sup>2</sup> 「知的財産推進計画 2015」 知的財産戦略本部（平成 27 年 6 月 19 日）

## 2. 調査の概要

特許庁が貸与する審決文 100 件について英語の人手翻訳文を作成し、作成した人手翻訳文を基準翻訳文として審決の機械翻訳文と比較することにより、機械翻訳文の精度や課題を調査・分析する。

また、本事業で作成した人手翻訳文をもとに、日英対訳コーパス及び日英機械翻訳用辞書を作成する。

### 2.1 実施体制

本事業に従事した担当者及び役割は下記の通りである。

責任者/担当者	作業内容	人数
統括責任者	本事業全体に対する統括を行う。	1 名
翻訳品質管理責任者	本事業の審決文の人手翻訳及び辞書データ登録語句の品質に対して責任を負う。	1 名
機械翻訳用辞書責任者	対訳コーパスと機械翻訳用辞書の作成において、フォーマットデータの作成と管理、作業者への指導・監督を行う。	1 名
進捗管理責任者	本事業全体の実務面の取りまとめとして、各責任者との連携、運用管理に対して責任を負う。	1 名
翻訳者	調査対象の審決文 100 件を人手により英訳する。	約 20 名
第一校閲者	英訳された審決文を英語表現の観点から校閲を行う。	約 6 名
第二校閲者	英訳された審決文に訳抜け、誤訳等がないか原文との突き合わせによる校閲を行う。	約 10 名
コーパス作成担当者	アライメントツールで機械抽出された日英対訳コーパスの確認・修正を行う。	約 4 名
調査・分析担当者	自動評価結果及び人手評価結果より集計データを作成し、各種機械翻訳文の精度比較や文単位の精度調査を行う。	約 11 名
辞書データ作成担当者	機械翻訳が誤訳となっている箇所や、辞書データの登録により機械翻訳の精度向上に効果が見込める語句を選定し、データ登録する。	約 6 名

表 2.1-1 実施体制



### (1) 基準翻訳文の作成

特許庁が指定した日本語の審決文 100 件を英語に人手翻訳し、その後第一校閲者及び第二校閲者による重畳的な校閲工程を経た人手翻訳文を作成した。

基準翻訳文の作成にあたっては、全 100 件の審決文について 10 件～13 件を 1 バッチとした計 9 バッチでバッチ単位に設定したスケジュールで進捗管理を行った。

審決文 100 件の四法毎、審判種別毎の内訳は以下のとおりである。

審判種別	四法			対象件数	対象件数の割合
	特許	商標	意匠		
無効	27	5	2	34	34.0%
不服	46	4	3	53	53.0%
判定	3	1	0	4	4.0%
訂正	0	0	0	0	0.0%
取消	0	1	0	1	1.0%
異議	6	2	0	8	8.0%
合計	82	13	5	100	100.0%

表 2.3-1 審決文 100 件の四法毎／審判種別毎件数

翻訳者・校閲者の採用条件：

#### (i) 翻訳者

産業財産権の出願手続等の事務手続に関する文書の日英翻訳経験を 1 年以上有する者

配置人数：約 20 名

#### (ii) 第一校閲者

英語を母国語とし、産業財産権の事務手続に携わる経験（弁理士としての業務経験等）

もしくはそれに足る知識を有する者

配置人数：約 6 名

#### (iii) 第二校閲者

産業財産権の事務手続に携わる経験もしくはそれに足る知識を有し、かつ、日英翻訳の校閲経験を 5 年以上有する者

配置人数：約 10 名

## (2) 英訳書誌記入済審決リストの作成

校閲された審決の英訳文から書誌事項を抽出したリストを作成した。

## (3) 日英対訳コーパスの作成

### (i) 審決対訳コーパスの作成

2.3(1)で作成した審決文 100 件の基準翻訳に基づき、原則として審決原文の日本語一文とそれに対応する英文を対応付けて対訳コーパスを作成した。

### (ii) 判決対訳コーパスの作成

特許庁が貸与する最高裁判所及び下級裁判所の判決の英訳文 1,856 ページを対象に、該当の判決原文を裁判所 HP から入手し、(i)と同様の形式で対訳コーパスを作成した。

## (4) 日英機械翻訳用辞書の作成

機械翻訳文の精度評価の結果と得られた知見をもとに、審決の機械翻訳文の精度向上に効果が見込める語句を選定し、辞書データ作成を行った。

審決文 1 件あたり 10 語程度（合計 1,000 語程度）となるよう辞書データを作成し、既存の辞書データとの重複を避けるため、作業者には原文に既存の辞書データが存在する語句に色づけを施したファイルを渡し、色づけされた語句は避けて辞書データ登録を行うよう指示した。

また、日本語データの切り出しを行う際に、専門用語抽出ツール（「言選」<sup>3</sup>）の使用に加え、自社開発の頻出語抽出ツールも使用した。複数のツールを使用することで登録後の検討漏れを防ぐとともに、辞書として登録すべき日本語範囲をより適切に切り出すことが可能となるため、両者の出力結果から最適なものを選定した。

なお、本事業で作成した辞書の種類はユーザ辞書である。これは、数字以外の可変部分を含まない語や句が対象となり、可変部分は数字のみ設定できるため「平成 26 年 4 月 25 日付け手続補正」のように、数字以外の訳を指定したい場合も本辞書に登録できる。

## (5) 審決汎用文例リストの更新

2.3(4)で作成した辞書のうち、汎用性が高いと考えられる用語について、リストに追加・更新を行った。また、更新したリストは、翻訳者及び校閲者間で随時共有し、英訳文の訳語表現の統一を図った。

---

<sup>3</sup> <http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb.html>

#### (6) 機械翻訳文の調査・分析

本事業で調査対象とした 100 件の審決文のうち、下記表に記載の 46 件の審決文を自動評価対象文とした。

審判種別	四法			対象件数	対象件数の割合
	特許	商標	意匠		
無効	11	1	1	13	28.3%
不服	23	2	1	26	56.5%
判定	1	0	0	1	2.2%
訂正	0	0	0	0	0.0%
取消	0	1	0	1	2.2%
異議	3	2	0	5	10.9%
合計	38	6	2	46	100.0%

表 2.3-2 四法別／審判種別別の評価対象件数

また、当該 46 件の審決文のうち、259 文を人手評価対象文として選定した。選定基準については「4. 審決の機械翻訳の人手評価」で述べることにする。

次に、「機械翻訳文の調査・分析」の際に用いた評価対象となる日英機械翻訳文について説明する。

#### (i) 機械翻訳の説明及び取得方法

本事業では、下記の表 2.3-3 に記載の 7 種類の機械翻訳を比較調査するため、評価対象 46 件（約 15,954 文）の基準翻訳文一文に対し、以下の 7 種類の機械翻訳文をそれぞれ取得した。

機械翻訳 No. <sup>4</sup>	翻訳サイト又はエンジ ンの種類	学習データ内容	機械翻訳の手法
① (み自①)	みんなの自動翻訳@ TexTra®	特許 NMT がベースのエンジン に、見解書 <sup>5</sup> コーパスデータ 1/2 程度を追加学習した	NMT
② (み自②)	みんなの自動翻訳@ TexTra®	特許 NMT がベースのエンジン に、見解書コーパスデータす べてを追加学習した	NMT
③ (み自③)	みんなの自動翻訳@ TexTra®	特許 NMT がベースのエンジン に、見解書コーパスデータす べてと過去作成分の審決コー パスデータ及び本事業で作成 した審決コーパスデータを追 加学習した	NMT
④ (み自④)	みんなの自動翻訳@ TexTra®	特許 NMT がベースのエンジン に、過去作成分の審決コーパ スデータ及び本事業で作成し た審決コーパスデータを追加 学習した	NMT
⑤ (J-Plat ⑤)	J-PlatPat 英語版	-	ルールベース
⑥-1 (SMT⑥-1)	特許庁から貸与された SMT エンジン	特許公報（ファミリー公報由 来の対訳コーパス）	SMT
⑥-2 (NMT⑥-2)	特許庁から貸与された NMT エンジン	特許公報（ファミリー公報由 来の対訳コーパス）	NMT

表 2.3-3 本事業で使用した機械翻訳 7 種

「み自①」～「み自④」の「みんなの自動翻訳@TexTra®」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サイト<sup>6</sup>で、ユーザ独自のコーパスを登録・学習させ、ユーザ独自の機械翻訳エンジンを作製することができる機能を提供している。

本事業では、このユーザ独自のコーパスを登録し訓練する機能（オリジナル自動翻訳

<sup>4</sup> カッコ内の表記を本文やグラフ等で使用する機械翻訳の名称とする。

<sup>5</sup> 見解書とは、審査官が発明の特許性（産業上の利用性、新規性、進歩性）についての肯定的又は否定的な見解を示す文書のこと。

<sup>6</sup> <https://mt-auto-minhon-mlt.ucri.jgn-x.jp/>

機能という。)を利用して、表 2.3-3 の①から④の「登録・学習したデータ内容」に記載の条件でそれぞれコーパスを登録・訓練した(アダプテーションという)。

「み自①」～「み自③」で追加学習用として登録した「見解書コーパスデータ」について、1,773,809 文対のコーパスデータから、「み自①」はデータの上半分(886,904 文対)、「み自②」及び「み自③」はすべてのデータをそれぞれ登録した。

見解書コーパスデータについて、特許庁貸与時のデータは2,477,997 文対だったが、以下の条件に当てはまる不完全なデータは除外としたため、学習データとして使用したのは1,773,809 文対となった。

- ・日本語文・英語文が同じで重複している。
- ・日本語と英語が1文対1文で対応していない。
- ・文の対応具合を示すスコア値がゼロ。
- ・文長が1文あたり1,000 バイト以上となるもの(学習エラーとなってしまうため)。

また、「み自③」及び「み自④」の「過去作成分の審決コーパスデータ及び本事業で作成した審決コーパスデータ」については、特許庁から貸与された過去3事業分の審決コーパスデータ(約55,833 万文対)と本事業で作成した15,075 文対(37 件分)の合計70,908 文対の審決コーパスデータを登録した。

加えて、「みんなの自動翻訳@TexTra®」では、汎用モデル、特許特化モデル、IT 特化モデル、特許請求項特化モデルといった分野に応じたエンジンを選択できる仕様になっており、本事業の「み自①」～「み自④」では、すべて「特許特化モデル」を選択して機械翻訳した英訳文を取得した。

「J-PlatPat⑤」の J-PlatPat 英語版は、以下の手順で J-PlatPat 英語版サイト<sup>7</sup>から、取得したい審判番号を入力し、「Results」で出てきた該当番号をクリックして表示された機械翻訳後の英訳文をテキスト抽出した。

---

<sup>7</sup> [https://www4.j-platpat.inpit.go.jp/eng/shinpan/spdb\\_en/SPDB\\_EN\\_GM101\\_Top.action](https://www4.j-platpat.inpit.go.jp/eng/shinpan/spdb_en/SPDB_EN_GM101_Top.action)



[Japanese](#) | [Top page](#) | [Help list](#) | [Site map](#) | [JPO](#) | [INPIT](#)

Help desk (9:00-21:00)  
 (+81)3-6666-8801  
[helpdesk@j-platpat.inpit.go.jp](mailto:helpdesk@j-platpat.inpit.go.jp)


 National Center for Industrial Property Information and Training

---

[Patent & Utility Model](#) | [Design](#) | [Trademark](#) | **[Trial & Appeal](#)**

[Top page](#) > Trial & Appeal > Publication DB of trial & appeal decisions

**Publication DB of trial & appeal decisions** [Help](#)
[Search](#) → [List](#) → [Detail](#)

You can make a reference about trial and appeal decisions (a decision of appeal for rejection, trial for invalidation, trial for correction, trial for rescission), a decision of opposition and a judgment out of various gazettes by their numbers.  
 \* NOTICES \*  
 These decisions and gazettes have been machine-translated into English and the translations may not be consistent with the original precisely. Especially a long sentence as seen in the Claim will reduce the translation accuracy.  
 JPO and INPIT are not responsible for any damages caused by the use of this translation.

Publication issued, and updates schedule, please refer to the [NEWS](#).

[Translation of selected decisions\(JPO site\)](#) | [About Trial/Appeals\(JPO site\)](#)

**Kind**  
 Published trial & appeal decision  
  Judgment gazette  
  Published trial & appeal decision / Judgment gazette

**Document Number**





[Japanese](#) | [Top page](#) | [Help list](#) | [Site map](#) | [JPO](#) | [INPIT](#)

Help desk (9:00-21:00)  
 (+81)3-6666-8801  
[helpdesk@j-platpat.inpit.go.jp](mailto:helpdesk@j-platpat.inpit.go.jp)


 National Center for Industrial Property Information and Training

---

[Patent & Utility Model](#) | [Design](#) | [Trademark](#) | **[Trial & Appeal](#)**

[Top page](#) > Trial & Appeal > Publication DB of trial & appeal decisions

**Publication DB of trial & appeal decisions** [Back](#) [Help](#)
[Search](#) → [List](#) → [Detail](#)

You can make a reference about trial and appeal decisions (a decision of appeal for rejection, trial for invalidation, trial for correction, trial for rescission), a decision of opposition and a judgment out of various gazettes by their numbers.

**Results**

Results 1 records.

No.	Document Number	
1	<a href="#">JP,2017-005823,J1</a>	<a href="#">This human translation is available</a>



**Publication DB of trial & appeal decisions** [← Back](#) [? Help](#) [Search](#) [→ List](#) [→ Detail](#)

You can make a reference about trial and appeal decisions(a decision of appeal for rejection, trial for invalidation, trial for correction, trial for rescission),a decision of opposition and a judgment out of various gazettes by their numbers.

**Selected Gazette** [← Previous Document](#) 1/1 [Next Document →](#)

**JP,2017-005823,J1** [Human Translation](#)

[Detail](#) [PDF](#)

**\* NOTICES \***

**JPO and INPIT are not responsible for any damages caused by the use of this translation.**

- This document has been translated by computer. So the translation may not reflect the original precisely. Especially a long sentence as seen in the Claim will reduce the translation accuracy.
- \*\*\*\* shows the word which can not be translated.
- In the drawings, any words are not translated.
- Information on [Name (in original language)] is the original description filed by applicant.

[Management number] No. 1336628  
 [Total volume number] No. 219  
 (19)[Publication country]Japan Patent Office (JP)  
 (12)[Kind of official gazette]Decision-in-trial-to-patent gazette  
 [Publication date]Heisei 30(2018) March 30 (2018.3.30)  
 [Classification] The appeal/trial decision of a Decision-of-Refusal appeal  
 [Appeal or trial number]Appeal No. 2017-5823 (P2017-5823/J1)  
 [Date of demand for trial]Heisei 29(2017) April 24 (2017.4.24)  
 [Fixed date] Heisei 30(2018) January 19 (2018.1.19)  
 [Classification of appeal and trial board decisions]  
 P18 .121-Z (G01B)  
 [Demandant]  
 [Name]National University Corporation Chiba University  
 [Representative]  
 [Patent Attorney]  
 [Name]Saegusa and Partners.  
 [Demandant]  
 [Name]KANSAI PAINT CO. LTD  
 [Representative]  
 [Patent Attorney]  
 [Name]Saegusa and Partners.  
 [Indication of case]  
 Patent Application No. 2012-75100 "thickness measurement method, measuring device, thickness change measuring method, and measuring device" the case of appeal against the examiner's decision of refusal [Heisei 25(2013) October 7 publication of unexamined application and JP,2013-205252,A] has resulted in the following appeal decision:

「SMT⑥-1」は、特許庁から貸与された SMT エンジンを用いて審決対訳コーパス内の原文を機械翻訳して英訳文を取得した。

「NMT⑥-2」も同じく、特許庁から貸与された NMT エンジンを用いて審決対訳コーパス内の原文を機械翻訳して英訳文を取得した。

「SMT⑥-1」及び「NMT⑥-2」には、ファミリー公報由来の対訳コーパスが学習データとして適用されている。

## (ii) 機械翻訳手法の説明

次に、本事業で使用する7種類の機械翻訳に使われている手法について説明する。

機械翻訳 No.	翻訳サイト又はエンジンの種類	機械翻訳の手法
み自①	みんなの自動翻訳@TexTra®	NMT(*)
み自②	みんなの自動翻訳@TexTra®	NMT(*)
み自③	みんなの自動翻訳@TexTra®	NMT(*)
み自④	みんなの自動翻訳@TexTra®	NMT(*)
J-Plat⑤	J-PlatPat 英語版	ルールベース
SMT⑥-1	特許庁から貸与された SMT エンジン	SMT
NMT⑥-2	特許庁から貸与された NMT エンジン	NMT

\*「み自①」～「み自④」は表 2.3-3 の通りコーパスの学習内容又は分量がそれぞれ異なる。

表 2.3-4 機械翻訳に使用されている手法

本事業では、表 2.3-4 に記載のとおり、ルールベース、SMT、NMT の3つの手法による機械翻訳を使って調査・分析を行うが、まずこの3つの手法の違いについて簡単に説明する。

機械翻訳の手法は大きく分けて「ルールベース機械翻訳」と「コーパスベース方式」に大別され、コーパスベース方式には、SMT、NMT がある。

### ルールベース機械翻訳 (Rule-Based machine translation, RBMT)

法規則に比重を置いた機械翻訳である。

言語ルールと文法ルールの組み合わせに加え、一般用語辞書を使用する。専門辞書でトレーニングすることで、正確な用語を用いた一貫性のある翻訳を実現できる。

### 統計的機械翻訳 (Statistical Machine Translation: SMT)

辞書／用語集 (コーパス) を解析し、統計、確率を基にした機械翻訳である。

単語、句、構文に基づいた対訳データから統計モデルを通して訳出される。

### ニューラル機械翻訳 (Neural Machine Translation: NMT)

1つの脳をモデルにしたマルチ処理デバイスである大規模なニューラルネットワーク、いわゆる「ディープラーニング」と呼ばれる人工知能(AI)に自動学習させる機械翻訳である。

### 3. 審決の機械翻訳の自動評価

「表 2. 3-2 四法別／審判種別別の評価対象件数」に示した 46 件の審決文について、「2. 3 (6) 機械翻訳文の調査・分析」に記載の方法にて下記の 7 種の機械翻訳を用意して自動評価を行った。

下記表中の「み自①」～「み自④」は、「みんなの自動翻訳@TexTra®」に対して 4 パターンの異なる学習データを適用したものである。適用した学習データの詳細は「8. 付録」に記載する。

機械翻訳 No.	翻訳サイト又はエンジンの種類	適用した学習データ範囲
み自①	みんなの自動翻訳@TexTra®	見解書 1/2 程度
み自②	みんなの自動翻訳@TexTra®	見解書すべて
み自③	みんなの自動翻訳@TexTra®	見解書すべて、過去事業の審決全件、及び現事業の審決 37 件
み自④	みんなの自動翻訳@TexTra®	過去事業の審決全件、及び現事業の審決 37 件
J-Plat⑤	J-PlatPat 英語版	-
SMT⑥-1	SMT エンジン	特許公報 (ファミリー公報由来の対訳コーパス)
NMT⑥-2	NMT エンジン	特許公報 (ファミリー公報由来の対訳コーパス)

表 3-1 自動評価対象の 7 種の機械翻訳

本事業の目的は、審決に特有の表現を適切に翻訳する機械翻訳手法を調査することであるため、審決文中で引用される公報部分は自動評価の対象から除外した。また、「手続きの経緯」、「理由」等の見出し語についても、すでに辞書が整備されており、機械翻訳で正しい訳語に置き換わることが多いため、同じく自動評価の対象からは除外することとした。

上記 7 種の機械翻訳の自動評価結果について以下に説明する。

#### 3.1 審決の機械翻訳の自動評価の結果及び比較

##### (1) 機械翻訳全体

機械翻訳 7 種に対して BLEU と RIBES の 2 つの手法による自動評価を適用し、BLEU、RIBES それぞれの各機械翻訳の評価数値を下記のように比較した。

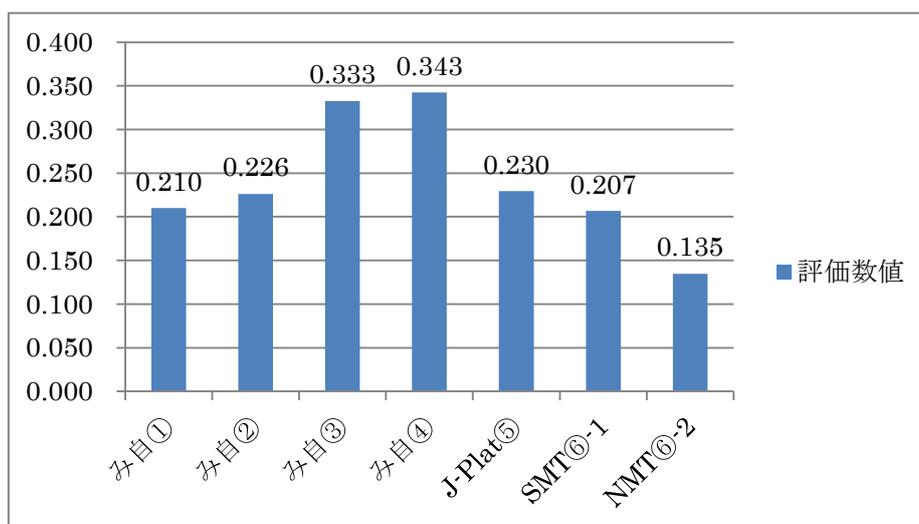


図 3.1-1 機械翻訳 7 種全体 (BLEU)

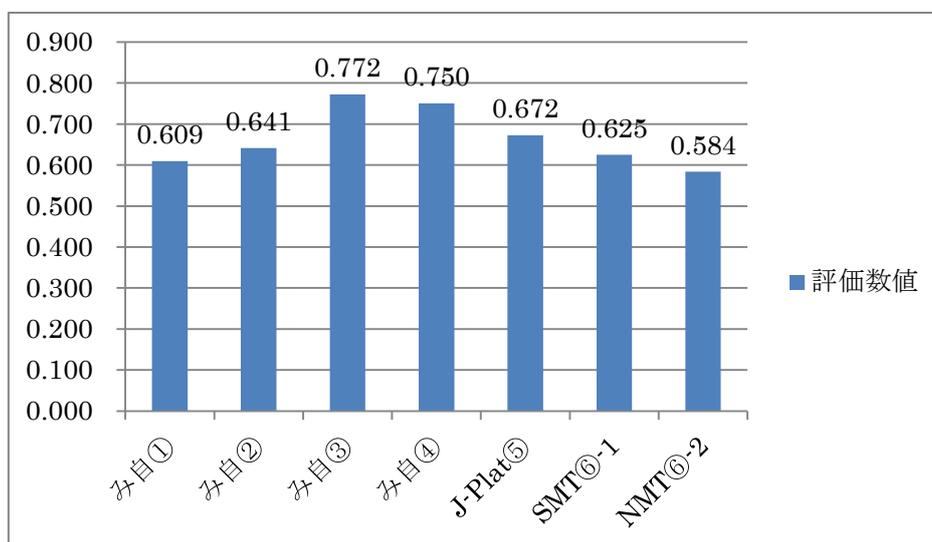


図 3.1-2 機械翻訳 7 種全体 (RIBES)

7 種の機械翻訳を比較したところ、BLEU では「み自④」（審決コーパスのみを学習させたもの）の評価がやや高く、RIBES では「み自③」（見解書すべてと審決コーパスを学習させたもの）の評価がやや高いという違いはあるものの、7 種の機械翻訳の中では、みんなの自動翻訳@TexTra®に一定量以上の対訳コーパスデータを学習させた機械翻訳の評価が比較的高いことが分かった。

また、評価数値が最も低いものは、BLEU、RIBES とともに「NMT⑥-2」（貸与エンジンのうちの NMT）であった。「NMT⑥-2」は特許公報を学習データとしたニューラル機械翻訳であることから、審決文の機械翻訳の品質を高めるには特許公報に加えて見解書や審決文といった審査関連文書も学習データに含める必要があることが分かる。

(2) 四法別

BLEU、RIBES それぞれの四法別の評価数値の結果を以下に示す。

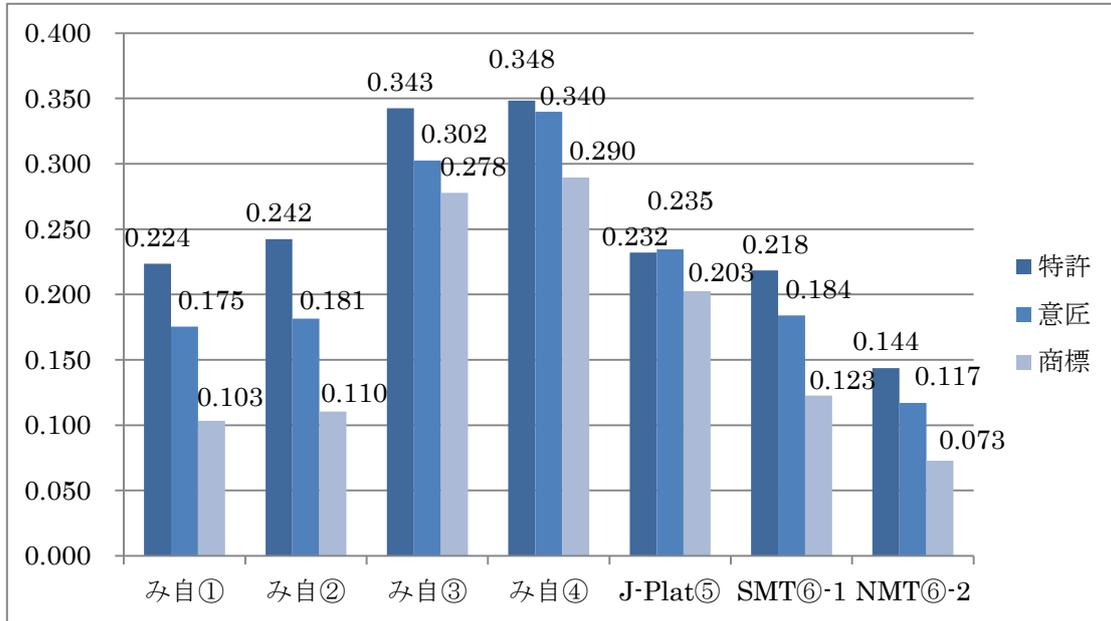


図 3.1-3 四法別 (BLEU)

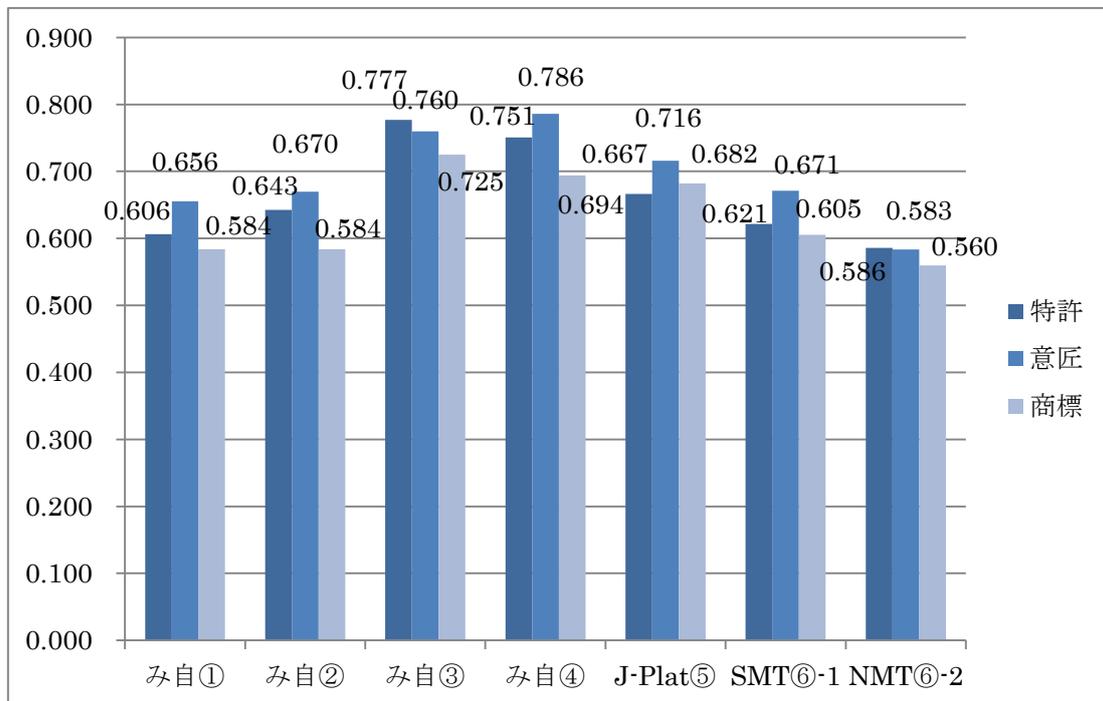


図 3.1-4 四法別 (RIBES)

四法別で見ると、BLEUでは「J-Plat⑤」を除くすべての機械翻訳において特許が最も高

い評価となった。RIBESでは7つの機械翻訳のうちの5つで意匠が最も高い評価となったが、対象とした46件の件数の内訳が特許38件、商標6件、意匠2件であるため、別の意匠の審決文においては異なる評価結果となる可能性もあり、意匠全体の傾向を十分に把握できているとは言えない。

BLEUの商標では、審決の対訳コーパスを学習していない「み自①」と「み自②」の評価数値が特許、意匠と比較して著しく低く、「み自③」と「み自④」において評価が大きく上がる結果となった。ここから、「み自③」と「み自④」で見解書だけでなく審決文を学習データに加えたことで、商標の精度が向上したことが分かる。

### (3) 審判種別別

BLEU、RIBESそれぞれの審判種別別の評価数値の結果を以下に示す。

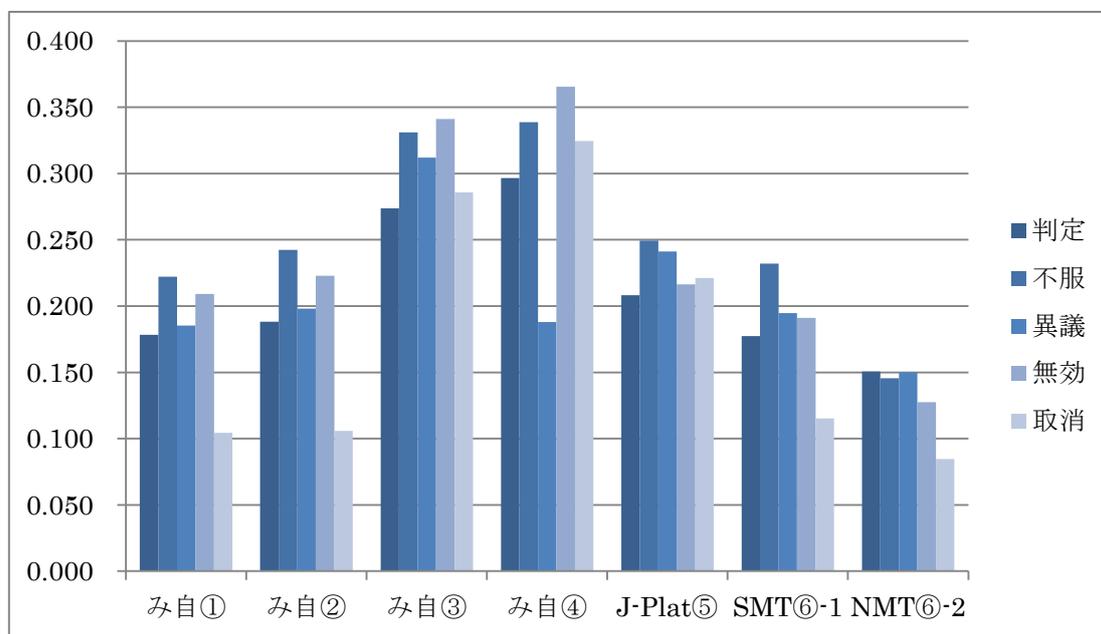


図 3.1-5 審判種別別 (BLEU)

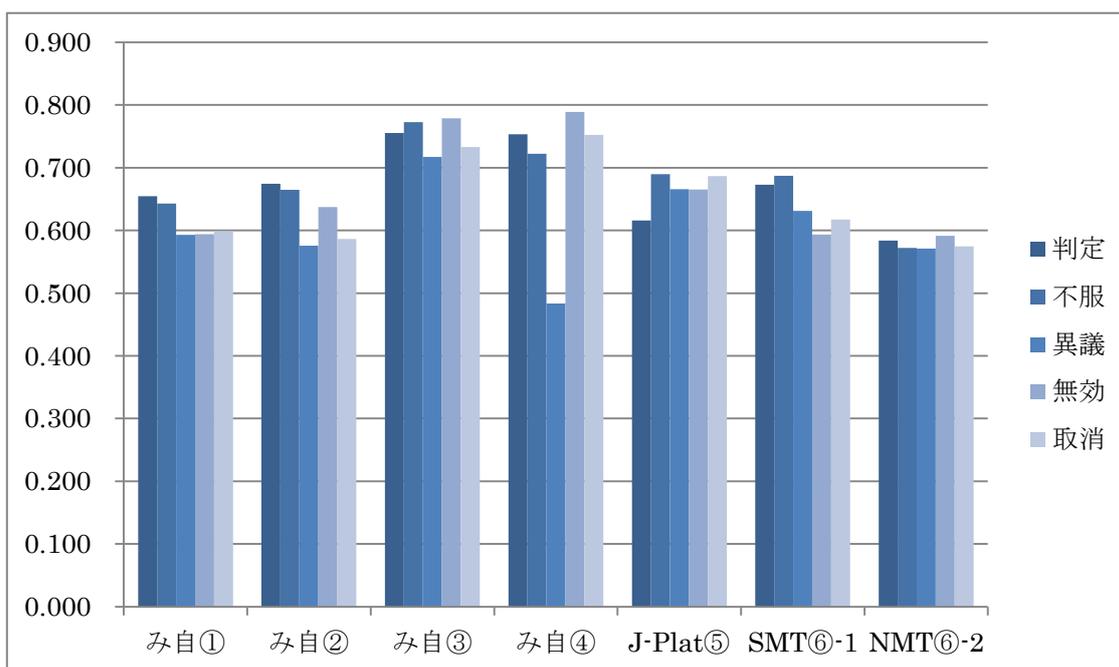


図 3.1-6 審判種別別 (RIBES)

審判種別別では、BLEU、RIBES とともに無効と不服の評価数値が高い結果となった。特に、BLEU における無効と不服の「み自②」から「み自③」への評価数値の上昇幅が顕著であるため、無効と不服の機械翻訳については審決の対訳コーパスを学習させたことが有効であったと考えることができる。

また、RIBES において「み自④」の異議が、他の 6 つの機械翻訳の中で著しく低い評価となっているが、ここから、今回評価対象とした 2 つの意匠については、見解書を学習データから外したことにより、異議の訳質を下げる結果となったと考えられる。

#### (4) 文の種類別

審決文の内容と各機械翻訳の精度との関係性を調べるために、審決文を記載内容に応じて下記の 6 つの種類に分けて各種類毎の自動評価数値を算出した。

種類名	定義
判断	結論を導き出すために審判官（合議体）が解釈した部分
認定	審判官（合議体）が認定した事実を列挙した部分
主張	当事者が主張した内容を説明した部分
手続の経緯	事件の手続きの経緯を説明した部分
証拠方法	引例等の証拠について記述した部分
結論	審決の結論部分

表 3.1-1 文の種類別の 6 項目

BLEU、RIBES それぞれの文の種類別の評価結果を以下に示す。

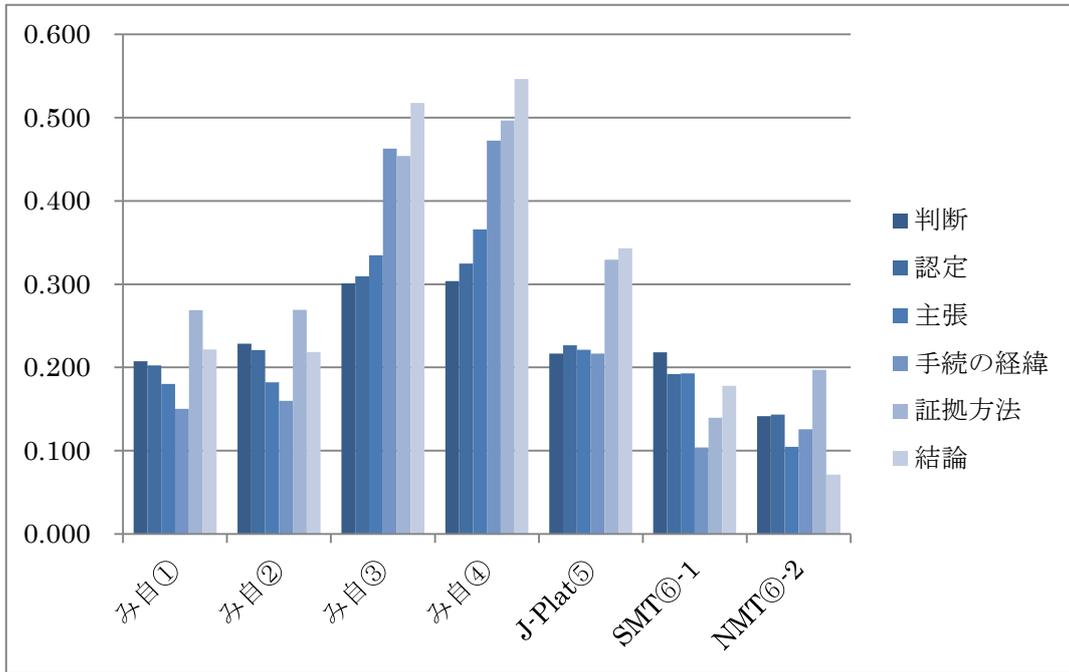


図 3.1-7 文の種類別 (BLEU)

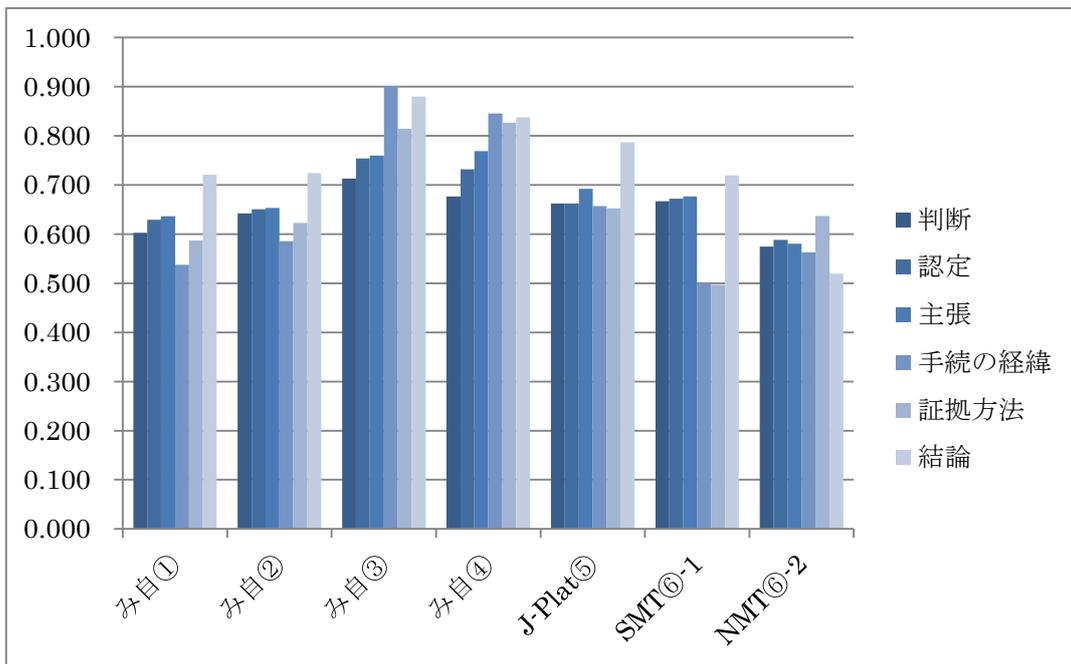


図 3.1-8 文の種類別 (RIBES)

BLEU において特に顕著であるが、結論と証拠方法は、学習データに審決の対訳コーパスが含まれる「み自③」と「み自④」の評価が高い結果となった。また、審決特有の表現が辞書データとして整備されている「J-Plat⑤」においても結論と証拠方法の評価数値が高

いことから、これらの種類については審決の対訳データの増強が機械翻訳の精度向上に非常に有効であるとみることができる。

なお、認定（審判官が認定した事実を列挙した部分）のうち、鉤括弧内に引用された公報部分は自動評価の対象から除外している。BLEU、REBES とともに認定の「J-Plat⑤」、「SMT⑥-1」、及び「NMT⑥-2」の評価値が比較的低めの値になっているのは、引用された公報部分を評価対象外としていることが影響したためと思われる。

### (5) 文長別<sup>8</sup>

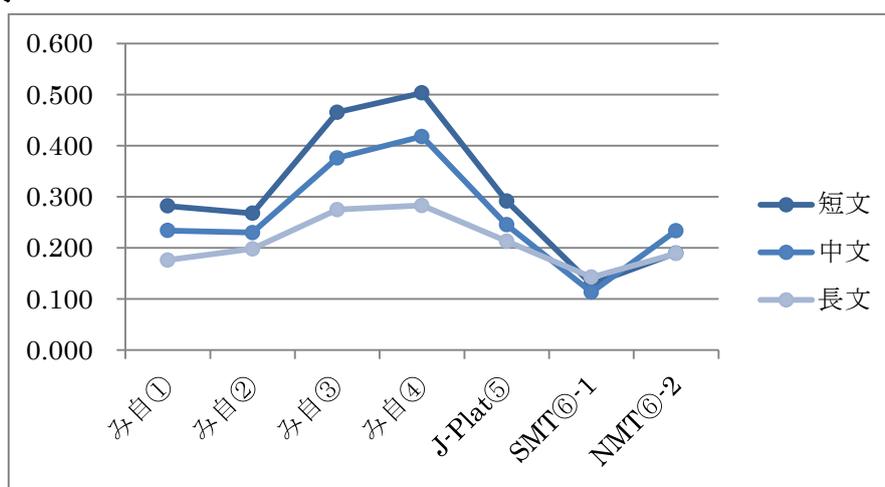


図 3.1-9 機械翻訳 7 種の文長別の比較 (BLEU)

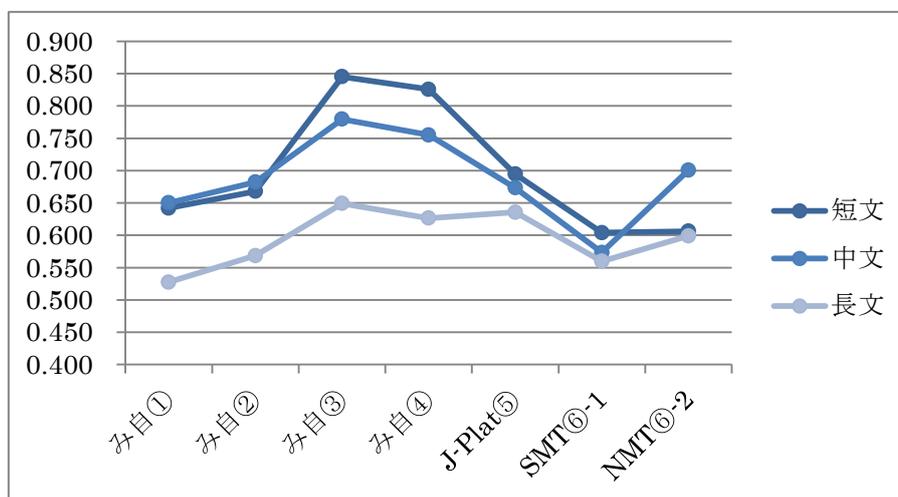


図 3.1-10 機械翻訳 7 種の文長別の比較 (RIBES)

BLEU、RIBES とともに短文、中文、長文の順に評価値が高くなっており、原文の日本語の文字数が短くなるほど機械翻訳の精度も向上することが明らかとなった。

<sup>8</sup> 評価対象の一文あたりの日本語文字数について集計を行い、一文の文字数により文を短文 (64 文字以下)、中文 (65-116 文字)、長文 (117 文字以上) の 3 区分とした。

また、短文と中文については、BLEU、RIBESともに「み自③」と「み自④」の評価が高い結果となった。このことから、短文と中文については学習データの分量を多くすることで機械翻訳の精度を高めることができることが分かる。

一方で、長文については BLEU、RIBESともに横ばいの折れ線になっているため、長文の機械翻訳の改善については学習データ以外の対策が必要になるものと思われる。

文長別の分析に関し、RIBESを用いてさらに下記の推移グラフを作成した。下記表は、四法別に文字数を降順にしてグラフ化し、四法毎の日本語原文の文字数の分布を示したものである。

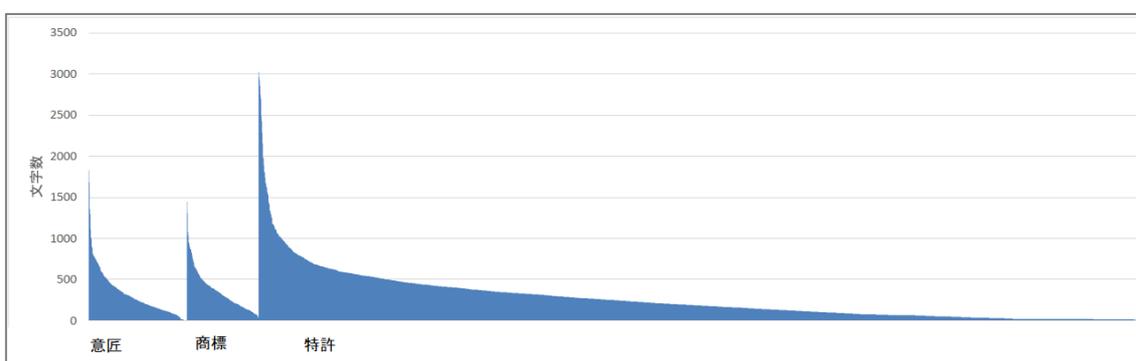


図 3.1-11 文長別の推移グラフ (RIBES)

意匠、商標に比べて特許にはより長い文が含まれるため、長文に起因する機械翻訳品質の強化を図る際は、意匠、商標よりも特許の難易度が高くなることが予想される。品質を強化する際は、意匠や商標と比べて長文になりがちな特許特有の条件を考慮した対応が必要になる。

また、上記グラフから、意匠や商標に比べて特許には短い文が高い割合で含まれることが分かる。特許において短めの文が多数含まれることが、「3.1 (2) 四法別」のグラフにおいて意匠、商標に比して特許の自動評価の数値が高くなった要因の1つと考えることができる。

#### 4. 審決の機械翻訳の人手評価

調査対象である審決文 100 件のうち、自動評価対象とした 46 件の各文から、書誌事項、見出し、技術的な説明文を除いた 259 文を人手評価文として選定した。

人手評価の対象とした 259 文の原文について、「み自③」、「J-Plat⑤」、「NMT⑥-2」の 3 種の機械翻訳からそれぞれ機械翻訳文を用意し、合計 777 文の機械翻訳文を対象に調査を行った。

機械翻訳 No.	翻訳サイト又はエンジンの種類	学習データ内容	機械翻訳の手法
み自③	みんなの自動翻訳@TexTra®	特許 NMT がベースのエンジンに、見解書コーパスデータすべてと過去作成分の審決コーパスデータ及び本事業で作成した審決コーパスデータを追加学習した	NMT
J-Plat⑤	J-PlatPat 英語版	-	ルールベース
NMT⑥-2	特許庁から貸与された NMT エンジン	特許公報（ファミリー公報由来の対訳コーパス）	NMT

表 4-1 人手評価文で調査した機械翻訳の種類

人手評価において下記の特徴を有する 3 種類の機械翻訳を用いることで、各機械翻訳の特徴及び傾向を把握することが可能になると考えられる。

「み自③」

現在、最も精度が高いと言われている NMT を使用しており、追加学習したデータ量が、「み自①」～「み自④」の中で最大の 1,844,717 文対である（審決と見解書を学習している。）。

「J-Plat⑤」

従来より使われているルールベース翻訳である。

「NMT⑥-2」

NMT を使用しているが、学習データは特許公報のみで、審決や審査書類は学習していない。

また、人手評価にあたっては、特許庁が公表している「特許文献機械翻訳の品質評価手

順 (ver. 1.0)」<sup>9</sup>に従って、人手により「内容の伝達レベル」の観点での評価を行い、内容伝達レベルを5段階に評価付けした（翻訳精度の最も高いものが評価値「5」、最も低いものが評価値「1」となる）。評価は3名で行い、その平均値を評価結果としている。

#### 4.1 審決の機械翻訳の人手評価の結果及び比較

##### (1) 審決全体

3種の機械翻訳を人手で評価した際の全体評価の結果を下記に示す。

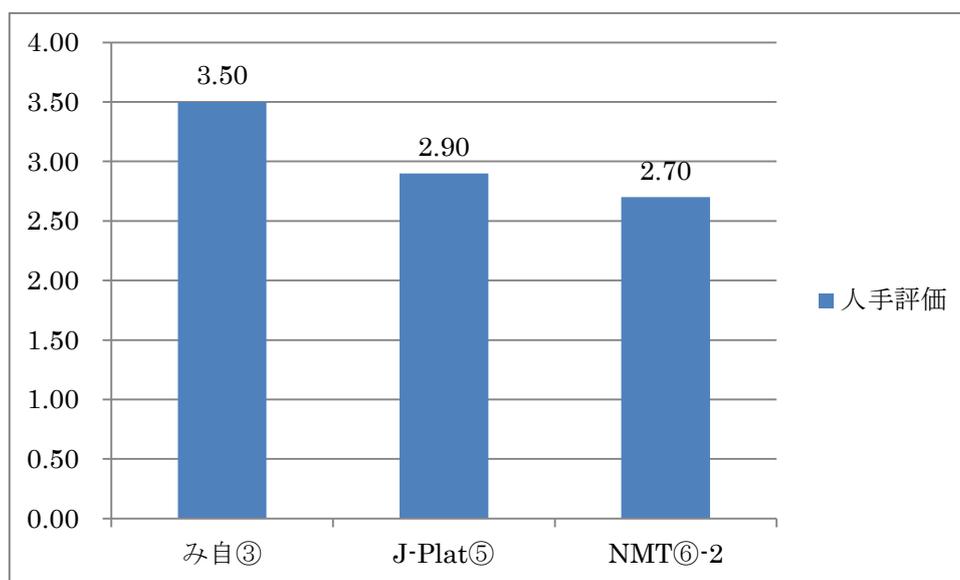


図 4.1-1 審決全体（人手評価）

3種の機械翻訳の人手評価の結果を比較したところ、「み自③」（見解書すべてと審決コーパスを学習させたもの）の評価が顕著に高い結果となった。「J-Plat⑤」は、「み自③」に次いで評価が高いが、もっとも評価の低い「NMT⑥-2」に値が近く、両者には大きな差はみられなかった。

この結果は、「図 3.1-1 機械翻訳7種全体 (BLEU)」及び「図 3.1-2 機械翻訳7種全体 (RIBES)」に示した自動評価とほぼ同一であり、人手で評価した結果においても審決と見解書を学習させた機械翻訳の精度がもっとも高いという結論となった。

##### (2) 四法別

次に特許、商標、意匠の四法別の結果を下記に示す。

<sup>9</sup> [https://www.jpo.go.jp/shiryou/toushin/chousa/pdf/tokkyohonyaku\\_hyouka/01.pdf](https://www.jpo.go.jp/shiryou/toushin/chousa/pdf/tokkyohonyaku_hyouka/01.pdf)

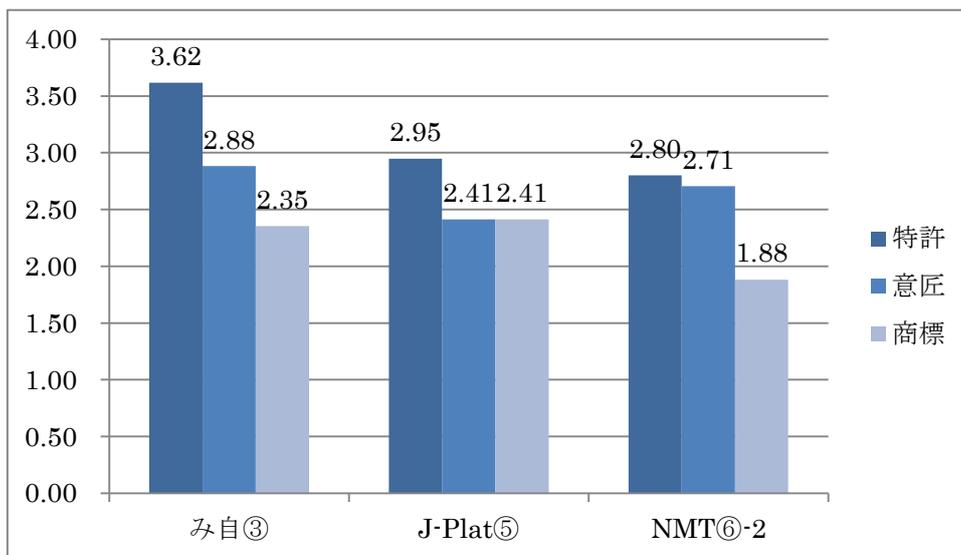


図 4. 1-2 四法別（人手評価）

四法別でみると、特許については「図 4. 1-1 審決全体（人手評価）」とほぼ変わらず、評価の高い順に「み自③」 > 「J-Plat⑤」 > 「NMT⑥-2」と並ぶ結果となった。

一方で、商標においては「J-Plat⑤」の評価が最も高く、意匠では「み自③」と「NMT⑥-2」の値がかなり近いものとなるため、特許に比べ、商標、意匠では審決と見解書の学習データの有効性はそれほど高くはないと考えることができる。

特許の「み自③」の評価が群を抜いて高くなっていることから、審決と見解書の学習データには特許の審決文に類似した表現が多く含まれているとみることができ、商標、意匠関連の審決文をより多く学習させることで、今後、商標、意匠の機械翻訳の精度も特許と同様に向上させることが可能とみることができる。

### (3) 審判種別別

次に、不服、無効、異議、取消の審判種別別の結果を下記に示す（判定は少数であるため除外するものとした）。

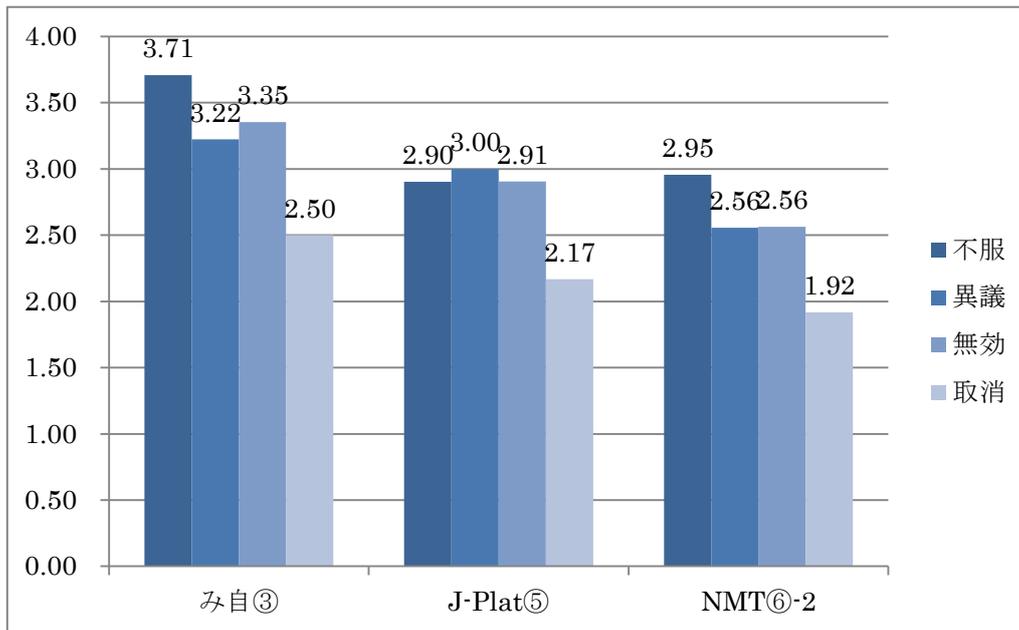


図 4.1-3 審判種別別（人手評価）

審判種別別では、不服、異議、無効、及び取消の4種すべてにおいて、「み自③」>「J-Plat⑤」>「NMT⑥-2」の順で「み自③」がもっとも高い評価となった。

ただし、取消は他3種と比べて全体的に値が低く、精度の向上については今後、検討が必要となるように思われる。

#### (4) 文の種類別

次に、認定、判断、主張、結論といった文の種類別の結果を下記に示す。

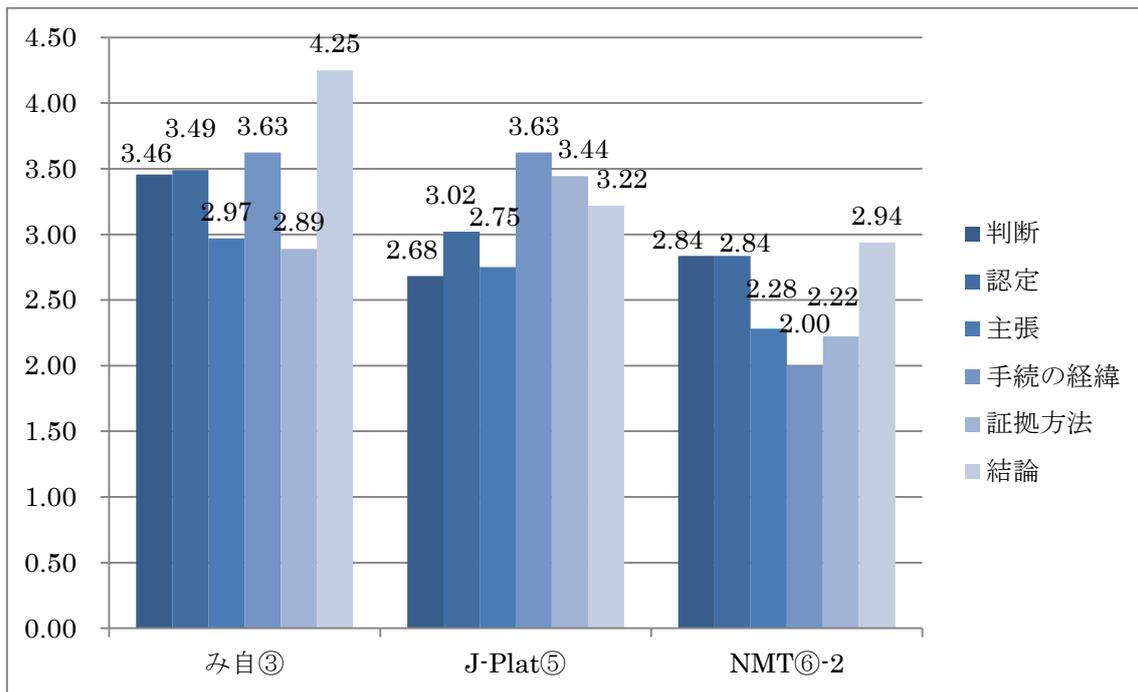


図 4.1-4 文の種類別（人手評価）

上記図は、「み自③」、「J-Plat⑤」、及び「NMT⑥-2」の3つの機械翻訳を、「表 3-1 自動評価対象の7種の機械翻訳」に示した審決文の記載内容別の6つのカテゴリーに分けて人手評価の結果を集計したものである。

全体を比べると、「み自③」の結論の値が4.25と群を抜いて高い結果となった。その次の階層をみると、評価値が3.5前後に位置するものに、「み自③」の判断、認定、手続きの経緯、「J-Plat⑤」の手続きの経緯、及び証拠方法がある。「NMT⑥-2」は最も値の高い結論であっても2.94と3を下回る水準となり、3つの機械翻訳の中ではすべてのカテゴリーにおいて評価が低い結果となった。

手続きの経緯、証拠方法、及び結論については「J-Plat⑤」の値が高く、特許用語が数多く登録されたルールベースの機械翻訳エンジンで、かなりの表現をカバーできているとみることができる。

#### 4.2 人手評価結果の分析コメント

人手評価結果について、「み自③」、「J-Plat⑤」、「NMT⑥-2」の3種の機械翻訳に出現した誤りを分析した。

その結果、NMTベースの「み自③」及び「NMT⑥-2」とルールベースの「J-Plat⑤」で頻出した誤りに特徴があることがわかった。まず「J-Plat⑤」は、文中の各用語や句の間の関係（係り受け）の誤りや誤訳が多く見られたが、重複訳や原文に存在しない語句が訳文に存在しているもの（湧き出し）、訳抜けはほとんど見られなかった。

また、「み自③」と「NMT⑥-2」は、重複訳や湧き出し、訳抜けが散見され、係り受けの誤り、誤訳、訳語不統一も見られた。

この「み自③」、「J-Plat⑤」、「NMT⑥-2」の機械翻訳に見られる翻訳上の誤りについて、特徴的なものを下記の7つの観点毎に具体例を挙げて説明する。

- (i) 湧き出し/重複訳
- (ii) 訳抜け
- (iii) 誤訳
- (iv) 訳語不統一
- (v) 係り受けの誤り
- (vi) 日英対訳コーパスの学習量に応じた機械翻訳精度の変化
- (vii) 日英対訳コーパスの登録による改善点や改悪点

(i) 湧き出し/重複訳

人手 評価 数値	原文	また、円形は下端側の1/6より上部より上部から上端部近傍まで直径を徐々に短くして直線状に設けられている。
	基準翻訳	Also, the circular shapes are linearly provided from the upper part above 1/6 on the lower end side to the vicinity of the upper end portion while gradually shortening diameters.
1	NMT⑥-2	In addition, the circular shape is provided in a straight line in a manner such as to gradually shorten the diameter from the top to the upper end of the upper end of the lower end side of the lower end side of the lower end of the lower end of the lower end of the lower end side than the upper end side of the lower end of the lower end side of the invention disclosed in document 1/6.
コメント (み自③) : 重複訳。「the lower end side of」が何度も出力されている。		

人手 評価 数値	原文	これに対して、請求人から平成25年7月29日に、上記請求項1～6に係る発明の特許を無効とすることについて、本件特許無効審判が請求され、平成26年3月4日付けで、被請求人から審判事件答弁書が提出された。
	基準翻訳	For this, the demandant filed a request for invalidation trial for the

		Patent on July 29, 2013 to invalidate the Patent for the inventions of claims 1 to 6. The demandee filed the written answer for trial on March 4, 2014.
2	み自③	On the other hand, regarding the invalidation of the patent of the inventions according to Claims 1 to 6 on July 29, 2013 on July 2014,, the trial regarding the invalidation of the patent invalidation trial was demanded on March 4,, and the demandee submitted a written reply for the trial case on March,.
コメント（み自③）：原文の後半に出てくる「平成26年3月4日」の訳が複数個所に分断され、なおかつ重複している。「7月」の訳語も重複している。		

(ii) 訳抜け

人手 評価 数値	原文	意願2016-7746「テレビジョン受像機」拒絶査定不服審判事件について、次のとおり審決する。
	基準翻訳	The case of appeal against the examiner's decision of refusal of Japanese Design Application No. 2016-7746, entitled "TELEVISION RECEIVER" has resulted in the following appeal decision.
1	NMT⑥-2	Claims 2016 to 7746 and "Television receiver" and "television receiver" are examined as follows.
コメント（NMT⑥-2）：「テレビジョン受像機」の訳が重複した結果「拒絶査定不服審判事件」の訳が抜けてしまった。また and で結ばれて主語が複数に訳されている。		

人手 評価 数値	原文	グラデーション模様を配した手摺については引用意匠1-1ないし4-2に見られるとおり、本件登録意匠の出願前より多数認められるものであるから、ガラス面板の上部の透明度を高く、下部の透明度を低く、中間の透明度をグラデーションで変化させたものは、その態様が既にありふれた態様といえるもので、本件登録意匠独自の特徴とはいえず、そこに格別の創意を認めることができないものである。
	基準翻訳	A number of the handrails with the gradation patterns, as shown in Cited Designs 1-1 to 4-2, are more likely to be recognized prior to the filing of the application for the Registered Design, and the form in which the glass face plate has high transparency in an upper part, low transparency in a lower part, and changes transparency in an intermediate part in a gradation manner can be said as an ordinary form, and thus it cannot be said that it is a unique characteristics of the Registered Design, and no particular creativeness is recognized

		in that.
3	み自③	As can be seen from Cited design 1-1 to 4-2, a person skilled in the art could easily conceive of a configuration in which the transparency of the upper part of the glass surface plate is high and the transparency of the lower part is reduced and the transparency of the middle is changed by changing the transparency of the lower part of the Registered design.
コメント（み自③）：赤字の原文の訳が抜けている。		

(iii) 誤訳

人手 評価 数値	原文	被請求人は、答弁書、口頭審理陳述要領書、平成26年7月10日付け上申書、平成26年10月27日付け上申書及び平成27年2月10日付け上申書において、以下の（h-1）から（h-5）の理由により、甲第1号証から「ラロキシフェンが、ラットの卵巣除去によって生じる骨密度の低下を有意に遅延した」との技術的事項を <b>把握できない</b> 旨主張している。
	基準翻訳	In the written answer, the Oral Proceedings Statement Brief, the written statement on July 10, 2014, the written statement on October 27, 2014, and the written statement on February 10, 2015, on the basis of the reasons (h-1) to (h-5) stated below, the demandee alleges that the technical matter "a decrease in bone density due to rat ovariectomy was significantly delayed by raloxifene" <b>cannot be found</b> in Exhibit A1.
2	J-Plat⑤	In a Written reply, an Oral proceedings statement brief, a Written Statement as of Heisei 26(2014) July 10, a Written Statement as of Heisei 26(2014) October 27, and a Written Statement as of Heisei 27(2015) February 10 a demandee, For the following Reasons of (h-1) to (h-5), from an Evidence A No. 1, a technical matter "raloxifene was intentionally delayed in deterioration of bone density occurred by spaying of a rat" <b>can be grasped</b> , and the purport opinion is twisted and carried out.
コメント（J-Plat⑤）：否定文が肯定文として訳出されている。		

人手 評	原文	また、 <b>単結晶からスライシングされたシリコンからなる基板</b> も本願の優先日前に既に周知である。
	基準翻	In addition, <b>a sliced silicon substrate from a single crystal</b> had

価 数 値	訳	been well-known before the priority date of the application, too.
4	み自③	In addition, a substrate comprising <u>silicon that has been sliced from a single crystal</u> was already well known before the priority date of the present application.
3	J-Plat ⑤	[ for example,, ] A substrate which consists of <u>silicon by which the slicing was carried out from a single crystal</u> is also already well-known before a priority date of this application.
<p>コメント：「単結晶からスライシングされたシリコン」の技術的な意味を理解していないと誤訳になってしまう箇所。「単結晶からスライシングされた」と「シリコンからなる」の両方が「基板」にかかる。</p> <p>(み自③) ⇒ 『シリコンがスライスされる』という意味になってしまっている。</p> <p>(J-Plat⑤) ⇒ 『シリコンにより』『スライシングがなされる』という意味が入ってしまっている。</p>		

人 手 評 価 数 値	原文	・本件訂正発明1は、「タモキシフェンより子宮癌のリスクの低い」の <b>に対し</b> 、引用発明は、この点についての記載がない点。(以下、「相違点2」という。)
	基準翻訳	- Corrected Invention 1 states that "the pharmaceutical formulation has lower risk of uterine cancer than tamoxifen" <b>but</b> Cited Invention does not include such a statement (hereinafter, referred to as "Different feature 2")
2	J-Plat⑤	- A point that a Cited Invention does not have a description about this point <b>to that</b> "in which a risk of a uterine cancer is lower than tomoxifen" as for the corrected invention of the case 1. (hereinafter referred to as " the different feature 2 ")
<p>コメント (J-Plat⑤)：原文の文意は「対し」で前後の内容を対照させており、基準翻訳のように but やあるいは whereas を使うべきところであろうが、訳文は「～に対する」で to が使われており、対照させる文意が汲み取れていない。</p>		

人 手 評 価 数	原文	また、請求人が意見書で提示した特開昭60-90899号公報に記載されているように、FZ法でシリコンからなる単結晶を製造する際に、酸素をドーピングする技術が周知であったにしろ、本願明細書の実施例は、 <b>【実施例】【0023】204mmの直径を有するシリコン単結晶を引き上げる場合、200mmより長い長さで無転位であ</b>
-----------------------	----	--

値		った
	基準翻訳	Therefore, even in the light of the description of the application, it is not understood that a silicon single crystal subject to the Invention is doped with oxygen. Furthermore, even if the technology for oxygen doping on the occasion of producing a silicon single crystal produced by FZ method was well-known as disclosed in Japanese Unexamined Patent Application Publication No. S60-90899 presented by the Appellant in the written opinion, an example described in the description of the present application is not doped with oxygen. Thus, it is obvious that oxygen doping is not carried out in the Invention even from the viewpoint of the above-mentioned well-known art. "[Examples] [0023] When a silicon single crystal having a diameter of 204 mm was pulled up, it was dislocation free in a length longer than 200 mm
2	み自③	In addition, as described in Japanese Unexamined Patent Application Publication No. 60 to 90899 published by the demandant in the written opinion, when producing a single crystal comprising silicon by the FZ method, the feature of doping oxygen was well known, but the examples in the description of the present application have a longer length than 200 mm when pulling up a silicon single crystal having a diameter of "[Examples] and paragraph [0023] 204 mm.
<p>コメント（み自③）：句点（。）で区切り対訳を作る場合、一文の中に引用箇所があると、（特に引用箇所に「。」が含まれる場合）、内容の切れ目が正確にならず、結果的に誤訳になる例。この原文をみると、この句点の後にさらに引用箇所の内容が続き、閉じ引用符の後に、この文章の述部があると思われる。しかし機械翻訳では引用箇所内の句点「。」を本文の句点とみて訳してしまっている。</p>		

(iv) 訳語不統一

人手評価数値	原文	本件訂正発明1の「ラロキシフェン」を表す化合物名が「～メタノン」であるのに対し、引用発明の「ケオキシフェン」を表す化合物名は、「～メタン塩酸塩」である点で、一応相違するが、ラロキシフェンとケオキシフェンが同じ化合物を意味することは当業者に明らかであるから（要すれば、THE MERCK INDEX THIRTEENTH EDITION, MERCKCO., INC. 2001, p.1452 「8190. Raloxifene」の項目参照）、甲第1号証の記載事項(エ)の「～メタン塩酸塩」は、「～メタノン塩酸塩」の誤記であり、
--------	----	--

		引用発明の「ケオキシフェン」は、本件訂正発明1のラロキシフェンの塩酸塩であると認められる。
	基準翻訳	The name of the compound representing "raloxifene" of Corrected Invention 1 is "--- methanone," whereas the name of the compound representing " <b>keoxifene</b> " of Cited Invention is "--- <b>methane hydrochloride</b> ." Thus, these compounds appear to be different. However, it is obvious to a person skilled in the art that raloxifene and keoxifene mean the same compound (if necessary, refer to THE MERCK INDEX THIRTEENTH EDITION, MERCKCO. , INC. 2001, p. 1452, the item "8190. Raloxifene"). "--- <b>methane hydrochloride</b> " of Described matter (D) of Exhibit A1 is error of "--- methanone hydrochloride" and " <b>keoxifene</b> " of Cited Invention is recognized as hydrochloride of raloxifene of Corrected Invention 1
3	み自③	The name of the compound name expressing "phenoxazine" of the corrected invention of the case 1 is "nano", whereas the name of the compound expressing " <b>Kenfenfen</b> " of the Cited invention is " <b>hydrochloride</b> "; however, it would be obvious to a person skilled in the art that benzylidene and <b>kenfen</b> are the same compound (if necessary, see the point of the MERCK INDEX THIVOLN EDITION, MERCKCO., INC. 2001, p. 1452 "8190. Roxazene"), it is recognized that the "- <b>methanesulfonate</b> " of the described matters in Evidence A No. 1 (D) is an error of "- metahydrochloride", and the " <b>kenfenfen</b> " of the Cited Invention is an error of the phenoxazine of the corrected invention 1 of the case.
コメント (み自③) : 訳語不統一。「ケオキシフェン」が Kenfenfen、kenfen、更に同審決文の中に kennofen と訳されているものもあった。「～メタン塩酸塩」についても同様に訳語が統一されていない。		

人手評価数値	原文	本件訂正発明2の「ラットにおいてタモキシフェンより子宮の上皮の高さの増大が少ないか、または、子宮の間質層への好酸球の浸潤が少ない」との記載に関して、発明の詳細な説明の記載を検討しても、「ラットにおいてタモキシフェンより子宮の上皮の高さの増大が少ないか、または、子宮の間質層への好酸球の浸潤が少ない」との条件を満たすために用量や、添加剤などを検討したとの記載はない。
	基準翻訳	Regarding the statement " <b>causing less increase in height of the epithelium of the uterus and lower eosinophil infiltration into the</b>

		<p>stromal layer of the uterus in rats, as compared with tamoxifen" in Corrected Invention 2, even considering the description of the Detailed Description of the Invention, there is no mention that the dosage and administration, additives, etc. were examined in order to satisfy the condition "causing less increase in height of the epithelium of the uterus and lower eosinophil infiltration into the stromal layer of the uterus in rats, as compared with tamoxifen."</p>
2	み自③	<p>Regarding the description of the Corrected invention of the case 2 that "the increase of the height of the child Miyanoé skin in rats or the invasion of eosinophils in the interstitial layer of the uteri is less than the description of the detailed description of the invention, it does not indicate that the amount or additive of the use of the case is examined in order to satisfy the condition that" the increase in the height of the tamoxifen child Miyanoé in the rat or the invasion of eosinophils in the interstitial layer of the uterine eye is less than that of the eye ".</p>
1	J-Plat ⑤	<p>The corrected invention of the case 2 "in a rat, increase of height of an epithelium of a uterus is less than tomoxifen, or being related with a description by or there being little permeation of eosinophile leucocyte to a stromata layer of a uterus" -- the detailed description of the invention, even if it considers a description, In order to fulfill conditions "increase of height of an epithelium of a uterus is less than tomoxifen in a rat, or there is little permeation of eosinophile leucocyte to a stromata layer of a uterus", there is neither a direction-for-use dosage nor a description that an additive agent etc. were examined.</p>
1	NMT ⑥ -2	<p>With regard to the wording "the increase in the height of the epithelium of the tamoxifen-child uterus in the rat in the rat is small or less than the infiltration of eosinophils into the stromal layer of the uterus", there is no indication that "the increase in the height of the epithelium of the tamoxifen-slave uterus in rats is small or the infiltration of eosinophils into the stromal layer of the uterus is small" in consideration of the disclosure in the Detailed Description of the Invention that "the increase in the height of the epithelial epithelium of the uterus is small or less" in the invention of the present invention (2).</p>

コメント：文中に同一文が2回引用されている文で、(み自③)、(J-Plat⑤)、(NMT⑥-2)の3種すべてにおいて2つの異なる機械翻訳がそれぞれ生成されてしまっている。また、(み自③)の前半部の引用文には、閉じ引用符(”)がない。

(v) 係り受けの誤り

人 手 評 価 数 値	原文	したがって、(a)～(d)の記載を総合すると、甲第1号証には、引用発明のラロキシフェンに、ヒトの閉経後の骨粗鬆症の動物モデルである卵巣切除ラットにおいて、骨粗鬆症の発症や進行を遅らせる効果がある、ことが記載されていた、といえる。
	基準翻訳	By summing up all these statements, Exhibit A1 may state that raloxifene of Cited Invention has an effect of delaying the onset and progression of osteoporosis in ovariectomized rats, which are an animal model of human postmenopausal osteoporosis.
2	J-Plat ⑤	Therefore, if a description of (a) - (d) is integrated, it can be told to an Evidence A No. 1 that what it is [ a thing ] effective in delaying the onset and advance of osteoporosis was described in raloxifene of a Cited Invention in an ovariectomy rat which is the animal model of osteoporosis after a human menopause.
コメント (J-Plat⑤)：「甲第1号証には、～ことが記載されていた」という係り受けが、「引用発明のラロキシフェンに、～ことが記載されていた」という係り受けになってしまっている。また、「ヒトの閉経後の骨粗鬆症の動物モデルである卵巣切除ラットにおいて、骨粗鬆症の発症や進行を遅らせる効果がある」というまとまりも分割されてしまっている。		

人 手 評 価 数 値	原文	記載事項(ア)、(イ)、(ウ)及び(ス)によれば、甲第1号証は、抗エストロゲン薬であるクロミフェンが、卵巣切除ラットの骨密度の低下を防止できることが知られていたところ、同じく抗エストロゲン薬であるケオキシフェンに着目し、高齢の卵巣切除ラットの骨密度への作用を確認することを、目的として行われた研究を開示した文献である。
	基準翻訳	According to Described matter (A), (B), (C), and (M), since antiestrogen clomiphene was known to be able to prevent a decrease in bone density of ovariectomized rats, Exhibit A1 is a document that focused on antiestrogen keoxifene and disclosed a study conducted aiming at confirming the effects of keoxifene on the bone density of elderly ovariectomized rats.



		described in <b>Evidence A No. 2</b> , the patent according to Claims 1 to 3 violates the provisions of Article 29(2) of the Patent Act, falls under Article 113(2) of the Patent Act, and should be invalidated.
4	み自③	In addition, the invention as in <b>claims 1 to 3</b> could be easily invented by a person skilled in the art from the invention described in <b>Evidence A No. 1</b> , and could have been easily invented by a person skilled in the art based on the invention described in <b>Evidence A No. 2</b> . Therefore, the patent according to Claims 1 to 3 was granted in violation of the provisions of Article 29 (2) of the Patent Act, and falls under Article 113 (2) of the Patent Act, and should be cancelled.
3	NMT⑥-2	In addition, the invention as in <b>claims 1 to 3</b> could have been easily invented by a person skilled in the art from the inventions disclosed in <b>claims 1 to</b> , and could easily be invented by a person skilled in the art from the inventions disclosed in <b>documents 2 to</b> ; therefore, the invention as in claims 1 to 3 is considered to have been violated by a person skilled in the art in violation of the provisions of Patent Article 29, paragraph [2] of the patent law, and should be cancelled. 2 113.
<p>コメント：(NMT⑥-2) は、特許公報のみを学習用データとしているため、審決や審査書類で頻出の「甲第○号証」の正確な訳出ができていない。「甲第○号証」を「claims ○」「documents ○」と訳しているため、特に前者は「請求項」の訳語と重複して文意を不明瞭にしている。それに対し、(み自③) は審決を多く学習しているため、該当語が正確に訳されている。</p>		

(vii) 日英対訳コーパスの登録による改善点や改悪点

改善点

人手評価数値	原文	したがって、 <b>本件補正</b> は、同法第159条第1項において読み替えて準用する同法第53条第1項の規定により却下すべきものである。
	基準翻訳	Accordingly, <b>the Amendment</b> shall be dismissed under the provisions of Article 53(1) of the same Act which is applied mutatis mutandis by replacing certain terms pursuant to

		the provisions of Article 159(1) of the same Act.
5	み自③	Therefore, <b>the Amendment</b> should be dismissed in accordance with the provisions of Article 53 (1) of the Patent Act which is applied mutatis mutandis in the provisions of Article 159 (1) of the Patent Act.
2	NMT⑥-2	Therefore, the <b>present invention</b> should be rejected in accordance with the provisions of PCT Article 53, paragraph [1], which is applied mutatis mutandis in the light of the same method as in claim 159 of <b>the present application</b> . 1.
コメント：(NMT⑥-2) では「本件補正」に対する訳が登録されていないと思われるが、(み自③) では基準翻訳と同じように正確な訳出がされており、改善が見られた。		

#### 改悪点

人手評価数値	原文	原査定の拒絶の理由に引用された引用文献1である特開2012-133471号公報には、「画像合成装置、画像合成プログラム、及び画像合成システム」(発明の名称) に関し、図面とともに次に掲げる事項が記載されている。
	基準翻訳	Japanese Unexamined Patent Application Publication No. 2012-133471 that is Cited Document 1 cited in the reasons for refusal of the examiner's decision describes the following matters with drawings <b>with regard to "Image Composer, Image Composition Program, and Image Composition System" (title of the invention).</b>
3	み自③	Japanese Unexamined Patent Application Publication No. 2005-133471, which is cited document 1 cited in the reasons for refusal stated in the examiner's decision, describes the following matters <b>relating to the "image synthesis device, image synthesis program, and image synthesis system" (title of the invention), with respect to the "image synthesis apparatus, image synthesis program, and image synthesis system" (title of the invention).</b> 2012.
2	NMT ⑥-2	Document 1 cited in the ISR cited in the ISR, which is cited in the ISR, <b>cites the following matter : "an image compositing device, an image synthesis program, and an image compositing system" (the name of the invention),</b> and the following features are disclosed in document 2012. 133471.

コメント（み自③）：青字部分が重複している。しかも重複している訳の間で、表現が異なっている（画像合成装置：image synthesis device と image synthesis apparatus）。また、湧き出しや訳抜けも見られた。

## 5. 自動評価を活用した審決の機械翻訳の分析

この章では、本事業の仕様書に指定された範囲を超えた独自のテーマに基づいて、BLEU、RIBES の自動評価手法を活用した審決の機械翻訳の多角的な分析を試みた。

2つのテーマに基づく分析について以下に記載する。

### 5.1 人手評価と自動評価との関係

「4. 審決の機械翻訳の人手評価」に記載したとおり、自動評価対象とした 46 件の審決文の各文から、書誌事項、見出し、技術的な説明文を除いた 259 文を対象に人手評価を行ったが、ここでは人手評価の対象とした 259 文について、人手評価と自動評価との相関関係を調査した。

下記のグラフは人手評価の対象とした 259 文の原文について、「み自③」、「J-Plat⑤」、「NMT⑥-2」の 3 種の機械翻訳からそれぞれ機械翻訳文を用意し、合計 777 文の機械翻訳文を対象に行った人手評価と BLEU、RIBES の 2 つの自動評価の結果を示したものである。

BLEU、RIBES の 2 つの自動評価の値は、対象とした 259 文の一文毎に評価値を測定して機械翻訳の種類毎にそれらの平均値を割り出して算出したものである。

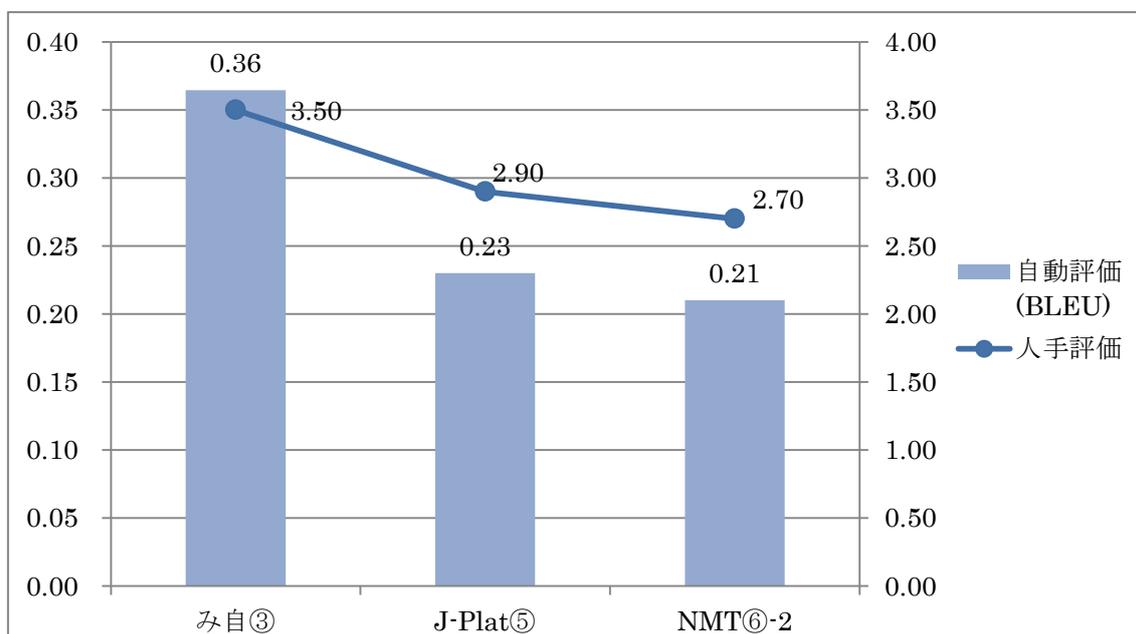


図 5.1-1 人手評価と自動評価との相関 (BLEU)

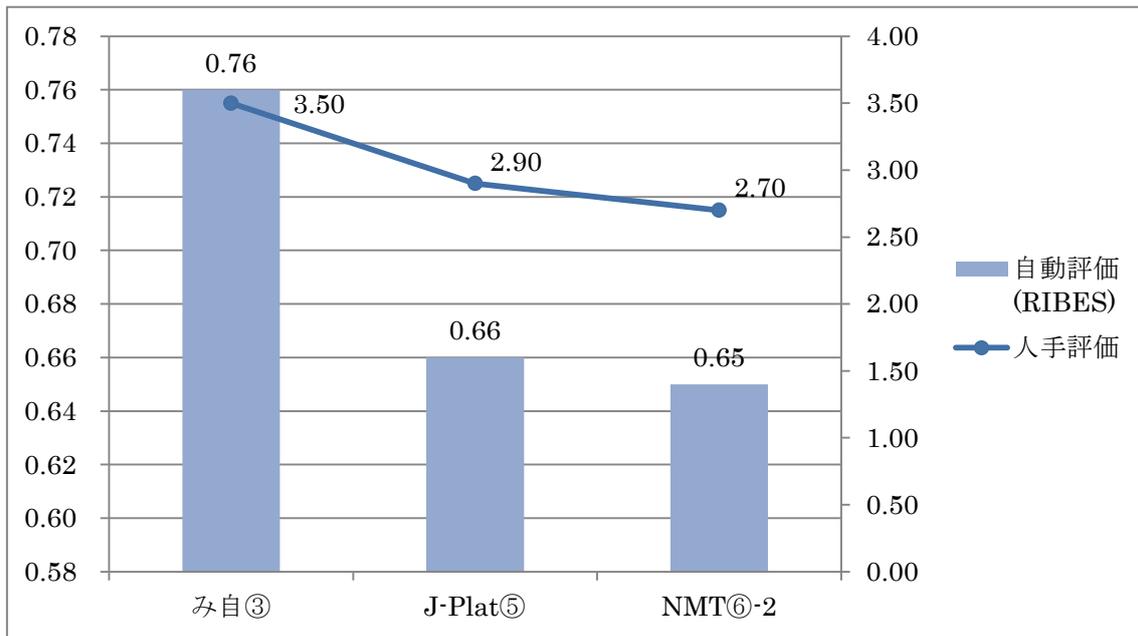


図 5. 1-2 人手評価と自動評価との相関 (RIBES)

上記の 2 つの図において、BLEU と RIBES それぞれの自動評価の値を棒グラフ、人手評価の値を折れ線グラフで示しているが、BLEU、RIBES のいずれにおいても自動評価と人手評価の値にきれいな相関関係がみられる結果となった。

また、順位についても BLEU、RIBES に大きな差はなく、どちらも「み自③」の評価が顕著に高く、「J-Plat⑤」と「NMT⑥-2」がほぼ同水準の値で続く結果となった。

今回の調査で人手評価を適用した原文文数はわずかに 259 文であり、「3. 審決の機械翻訳の自動評価」に示したように、大多数は自動評価手法を用いて機械翻訳の精度調査を行ったものであったが、上記の図のように人手評価と自動評価の間に明らかな相関関係がみられたことにより、自動評価による本調査の分析結果が信頼できるものであると言える。

## 5.2 審決文 100 件を対象とした分析

審決文 100 件を対象に、「NMT⑥-2」による機械翻訳文を作成し、審決文の機械翻訳の傾向や特徴について調査した。

みんなの自動翻訳@TexTra®を用いた「み自①」～「み自④」は審決文 100 件の半数近くを学習データとして用いて作成した機械翻訳であるため、審決文 100 件の機械翻訳の傾向や特徴を調べるには不適切と判断し、100 件全件の機械翻訳を同一条件で作成できる点を考慮して、本調査には特許公報のみを学習させた「NMT⑥-2」を採用することとした。

BLEU、RIBES を用いた自動評価結果を以下に示す。

(1) 審決文 100 件全体

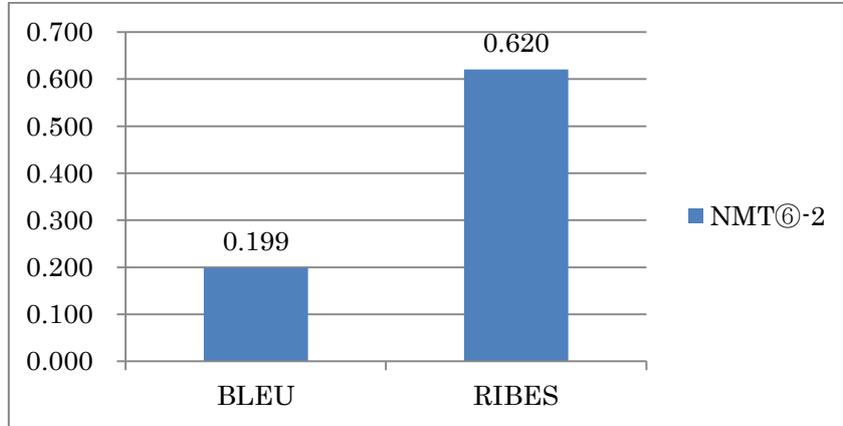


図 5. 2-1 審決文 100 件全体 (BLEU/RIBES)

審決文 100 件全体を BLEU と RIBES にかけてところ、上記の図のように BLEU が 0. 199、RIBES が 0. 620 という値になった。「3. 審決の機械翻訳の自動評価」に示した図をみる際に、これらの値を標準値と捉えて各観点に基づく評価の比較結果を考察することができる。

(2) 四法別

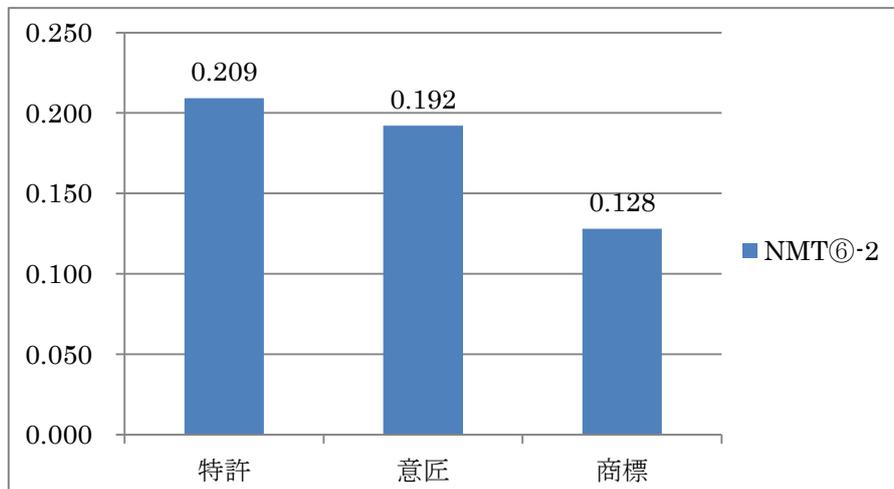


図 5. 2-2 四法別 (BLEU)

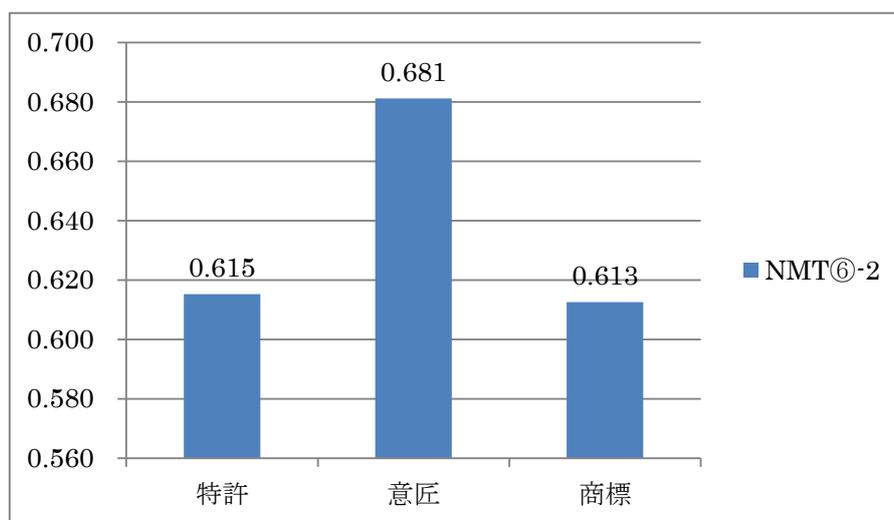


図 5.2-3 四法別 (RIBES)

審決文 100 件全体を四法別にみると、BLEU と RIBES とで評価が大きく異なる結果となった。

BLEU においては特許の評価が最も高いが、RIBES では意匠が最も高い値となった。

RIBES は、人手による訳と機械翻訳による訳とを比較する際に文の語順をより大きく考慮するものであるため、図 5.2-2 と図 5.2-3 の結果の相違に訳文の語順が影響したと推測することもできるが、明確な要因は定かではない。

### (3) 審判種別別

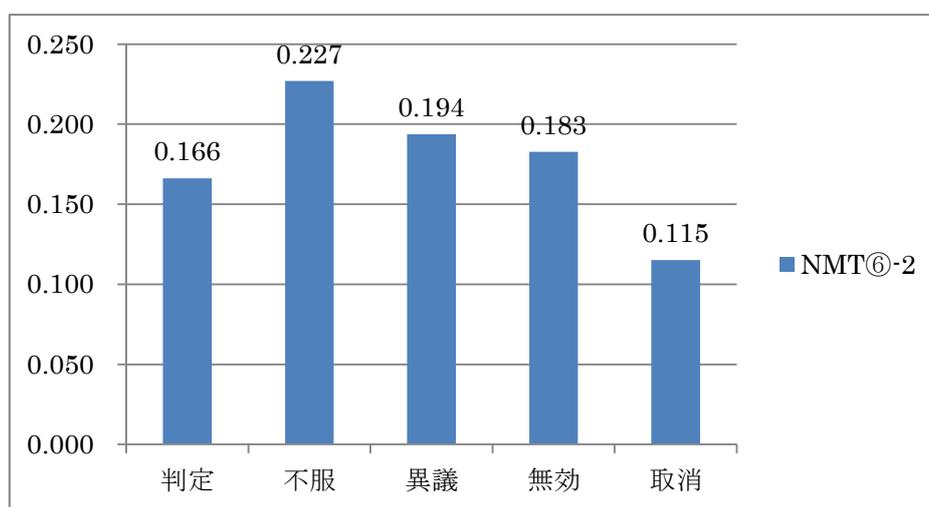


図 5.2-4 審判種別別 (BLEU)

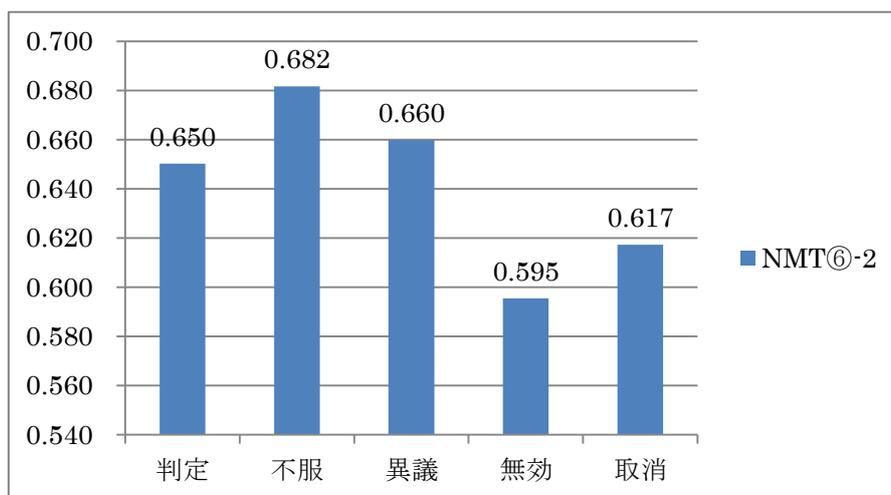


図 5.2-5 審判種別別 (RIBES)

審決文 100 件を審判種別別に比較したところ、BLEU と RIBES とともに不服、判定、及び異議の評価が高くなった。

無効のみ BLEU での評価が高く、RIBES では最も低い値になるといった違いが見受けられたが、無効を除く 4 種の順位については BLEU と RIBES とに大きな差はみられず、共通して取消の値が低い結果となった。

#### (4) 文の種類別

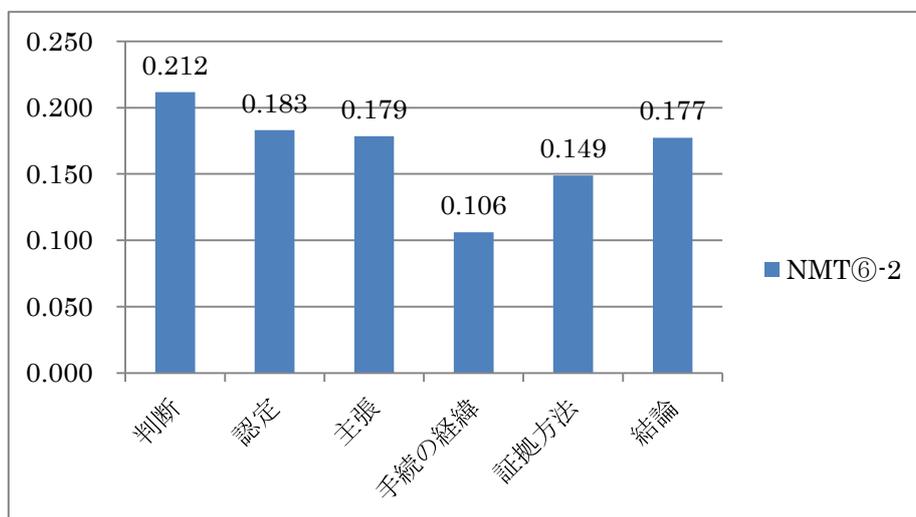


図 5.2-6 文の種類別 (BLEU)

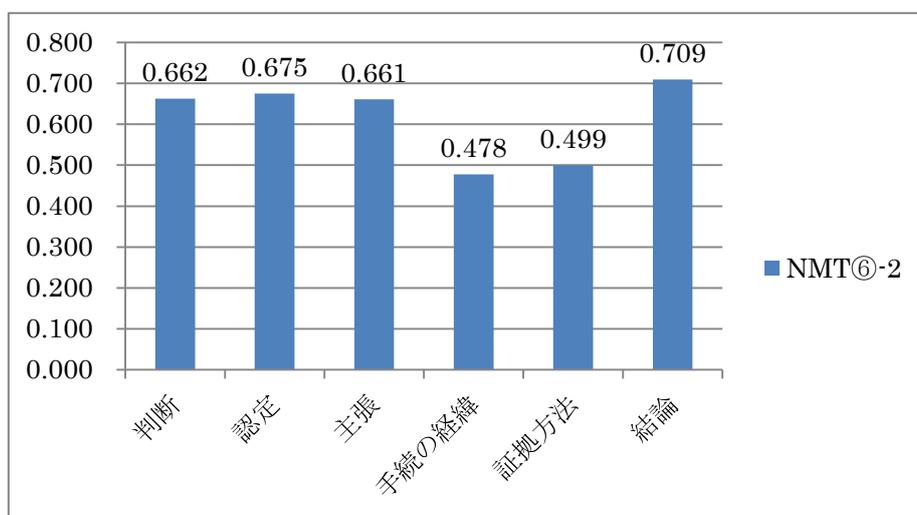


図 5.2-7 文の種類別 (RIBES)

審決文 100 件を文の種類別に比較したところ、BLEU と RIBES とともに判断、認定、主張、及び結論の評価が高い結果となった。

一方で、手続きの経緯と証拠方法は BLEU と RIBES とともに評価の値が低かった。「図 3.1-7 文の種類別 (BLEU)」や「図 3.1-8 文の種類別 (RIBES)」で示した結果と比較すると、手続きの経緯と証拠方法は特許公報を学習させた「NMT⑥-2」では適切な訳になりにくいものと思われる。「3.1 (4) 文の種類別」に記載した調査結果を考慮すると、見解書や審決文を学習データとすることで、手続きの経緯と証拠方法については今後、大幅な改善がみられるものと考えられることができる。

## 6. 辞書データの作成・分析

審決文 100 件を対象に行った辞書データの作成作業について説明する。

本事業では、審決文 1 件当たり 10 語程度の名詞句を原文、訳文の対訳形式で登録し、全体で 1025 語からなる辞書データを作成した。

### 6.1 登録語の編集における工夫

#### (1) 登録作業時の留意点

機械翻訳の精度向上に有効なデータとなるよう、辞書データの登録候補語を登録する際は下記の点に注意して作業を行った。

- ・既に登録された語は除外した

「2.3 調査方法」の「(4) 日英機械翻訳用辞書の作成」に記載の方法で辞書データを作成するにあたり、既存辞書内に含まれる語については審決文の原文中に色付けを行い、重複して登録しないよう注意した。

- ・審決特有の表現を優先的に登録した

上記のように既存辞書に該当する部分に色を付けたところ、審決特有の表現の中で色が付かない語句があまりなく、大半の名詞句は既に審決の辞書データに登録されている状態であったが、鉤括弧内の公報引用箇所はできるだけ登録対象とせず、審決特有の語を優先的に登録するようにした。

- ・短めの語句は避けて登録した

例えば、「申請」で登録すると「申請人」が適切に置換されないことがあるため、複合語を中心に極力長めの語句を優先して登録するようにした。

- ・多義語は避けて登録した

多義語を登録すると文脈によっては誤訳となることがあるため、一義的に訳が定まる訳語を中心に登録するよう注意した。

- ・冠詞、単複に注意して登録した

対象語の用例を考慮し、常に複数形で用いられるものは複数形で登録する等、冠詞や単複の形式に注意して登録した。

#### (2) 辞書データ作成時の留意点

今回、登録候補語を登録するにあたっては、審決に特有の表現をより多く登録できるよう、審決文毎に基準翻訳文と原文とを照合して原文全体の流れを踏まえた上で作業を進めるものとした。

登録すべき候補語が多数含まれる審決文においては適宜、機械翻訳文中の対訳を参照し、誤訳となるものに限定して登録する等の選定作業を行ったが、上述したように、1 件の審決文中の公報引用箇所を除く範囲において、既存辞書に登録されていない名詞句がごくわず

かであったため、既存辞書にない審決特有の表現となる名詞句は極力すべて登録することとした。

そのような方針で作成したため、「辞書登録語・登録候補語リスト」のフォーマット内の「派生番号」、「品詞種別」、「活用語の位置」、「誤訳箇所」の 4 項目については該当するものがなく、該当欄は空欄とした。

## 6.2 辞書データ作成についての提言

今後、より有効な辞書データを作成するための提言について以下に記載する。

### 既存辞書の整備

動詞や形容詞等、活用形が存在するものは適用された文脈によっては不適切な訳語となるため、使用前に十分に精査する必要がある。

また、名詞であっても語数が短いものは同様に文脈によって不適切な訳語となることがあるため、語数が少ないものは使用しないことを推奨する。

### 登録前のシュミレーションの実施

登録した辞書データを用いて機械翻訳すると「the the patent invention」と冠詞が二重化してしまう等の問題が発生することもあるため、サンプルデータを用いてテストし、登録候補語を使用前に十分に精査することを推奨する。

## 7. 審決の日英機械翻訳文提供に向けての課題

審決の日英機械翻訳文の作成における現状の課題と今後に向けた提言について以下に記載する。

### 7.1 審決の日英機械翻訳文の現状

現状の機械翻訳の問題点を調べるために、3つの審決文を対象に、みんなの自動翻訳@TexTra®の「み自③」と「み自④」との違いに焦点を絞って用語、フレーズ単位での誤訳調査を行った。その結果検出された機械翻訳の誤訳事例を下記に記載する。

#### 例1) 特許 不服 2017-5823

No.	日本語	基準翻訳文	「み自③」	「み自④」
1	～である点で共通する	is common to ... in a point of ~	is common in point of ~	is common in the point that ~
	「み自③」は「in point of」と"point"の前に冠詞がない。			
2	～についての判断	Judgment regarding ~	Examination on ~	Examination on ~
	judgement regarding~:~についての判断、 examination on~:~についての検討、考察、調査			
3	本願発明の構成を備えるようになる	comes to have the constitution of the Invention	would be provided with the configuration of the invention	includes the configuration <u>of</u> the Invention <u>of</u> the Invention
	「み自④」は"of the Invention"が重複している。			
4	当を得ない。	is improper	cannot be accepted	is groundless
	improper : 妥当でない、 cannot be accepted : 受け入れられない、 groundless : 根拠のない			
5	他の請求項に係る発明について審理するまでもなく	without examining the inventions <u>according to</u> other claims	without examining the invention <u>as in</u> the other claims	without examining the inventions <u>relating to</u> the other claim
	according to~ : ~次第で、~によれば、~に従って、 as in~ : ~と同様に、 relating to~ : ~に関して			

例 2) 特許 不服 2014-7822

No.	日本語	基準翻訳文	「み自③」	「み自④」
1	特許法 17 条の 2 第 5 項 2 号	Article 17-2(5)(ii) of the Patent Act	Article 17 (2) (I) of the Patent Act Article 5 (2) of the Patent Act.	Article 17 (2) (I) of the Patent Act … Claim 5
	「み自③」、「み自④」では「特許法 17 条第 2 項第 1 号」の意味になっており、湧き出し語もある。			
2	特許法第 17 条 の 2 第 5 項の規 定に違反するの で、	violates the provisions of Article 17-2(5) of the Patent Act	violates the provisions of Article 17 (2) (iii) of the Patent Act	violates the provisions of Article 17 (2) of the Patent Act
	「み自③」、「み自④」では「の 2」が第 2 項になり、第 5 項が訳抜け。			
3	更に…を備える (対象外)	, further comprising:	, which is provided with	, further comprising
	「み自③」は「更に」が訳抜け。			
4	両者は相当関係 にある。	the two are in a correspondence relation.	the two are in correspondence with each other, <u>the two are in correspondence with each other.</u>	the both are in a corresponding relation.
	「み自③」は訳語が重複している。			
5	において読み替 えて準用する	is applied mutatis mutandis by replacing certain terms pursuant	is applied mutatis mutandis by replacing	is applied mutatis mutandis pursuant to
	「み自④」は「読み替えて」が訳抜け。			

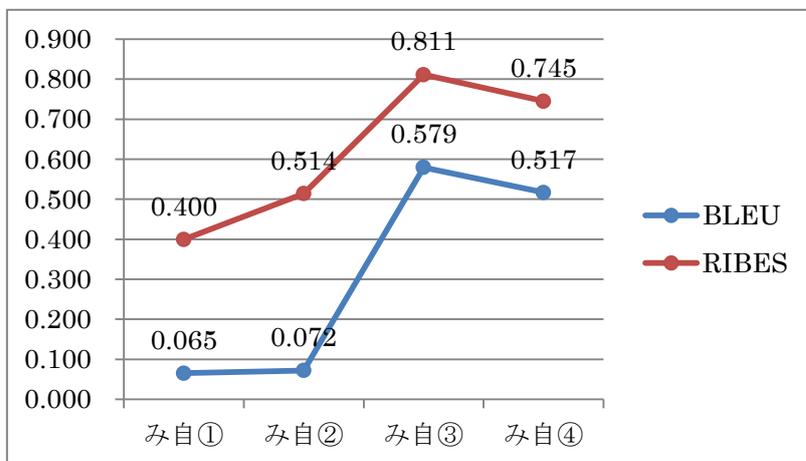
例3) 特許 無効 2012-800199

No.	原文	基準翻訳文	「み自③」	「み自④」
1	上記結論と同旨。	The gist is the same as the above conclusion.	The same conclusion as the conclusion above	The same as the above conclusion.
	「み自③」、「み自④」では、「上記結論と同じ」であり、「同旨」とは異なる意味になっている			
2	特願2005-88334	Japanese Patent Application No. 2005-88334	Japanese Patent Application No. 2013-88334	Japanese Patent Application No. 88334,
	「み自③」では、出願年の「2005」が「2013」になっている。「み自④」では、出願年が訳抜けしている。			
3	優先権主張：平成16年3月30日。	priority claimed on the basis of earlier application: March 30, 2004,	the priority claim based on the prior application : <u>2004 March 2005</u> ,...the earlier application on March 30,	the priority claim based on the earlier application : March 30, 2004,
	「み自③」では、年を複数含む湧き出し文が出現し、後続の文は年が抜けている。			
4	平成23年7月8日、設定の登録を受けた。	registered on July 8, 2011.	filed on March 8, 2011.	was granted a registration of the establishment on July 8, 2011. <u>The establishment of the establishment of the patent was registered on July, 2005.</u>
	「み自③」は月の誤訳があり、「み自④」では湧き出し文が出現し語句の重複と年の誤訳がある。			
5	請求人代理人、被請求人代理人の出頭のもと、	in the presence of the demandant's representative and the demandee's representative,	on the basis of the demandant's attorney and the demandee's allegation,	by the demandant's own agent, the demandee's own company,
	「み自③」は「請求人側弁護士と被請求人の主張に基づき」の意味に訳されている。			

	「み自④」も「請求人代理人と被請求人自身の会社により」と訳されている。			
6	以後本件は書面審理とするとされた。	were advised of the case being transferred to the examination by documentary proceeding.	and the following trial shall be made as a trial proceeding.	and the demandee assumed to be a document for trial later.
	基準翻訳文の訳文は、特許無効の審判が特許法第145条第1項の規定に基づき口頭審理で始まり、2回目以降は、通常審判長の「助言」（法律上は職権だが、両当事者は通常助言に同意する）により、書面審理に移行することを的確に表現している。しかし「み自③」、「み自④」とも原文の意味を殆ど表現できていない。			
7	その謄本は8月8日、両当事者に送達された。	copies of which were sent to both parties on August 8.	<u>Corrected book of the case was delivered to August 8, 2013 on August, 2013.</u>	delivered the certified copy to the both parties on August 8, 2013, and delivered to the both parties <u>on August 30, 2013.</u>
	「み自③」は「謄本」を誤訳し、日付に湧き出し語句を含む。「み自④」は日付関連の湧き出し文を含む。			
8	平成25年9月4日、一次審決を不服として知的財産高等裁判所に訴えを提起した	filed a suit on September 4, 2013 with the IP High Court against the first trial decision to complain about the first trial decision	had made the intellectual property high court as an unfair trial decision on September 4, 2013.	filed a suit on the Intellectual Property High Court with a first trial decision of September 4, 2013.
	「み自③」の訳は意味を成していない。「平成25年9月4日」が「み自④」では提訴した日付ではなく一次審決の日付になっていた。			

次に、みんなの自動翻訳@TexTra®の「み自①」～「み自④」を用いて学習コーパスデータを増減による審決の日英機械翻訳文の変化を調べることで、現状の審決の日英機械翻訳文の傾向を明らかにするよう試みた。

学習コーパスを増やしたことで訳質が改善した例  
例)

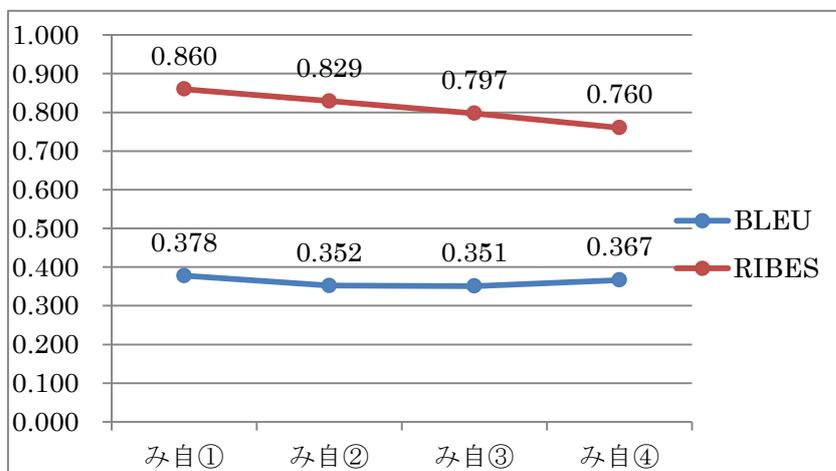


原文	よって、本件特許明細書の記載は、特許法第36条第4項第1号に規定する要件を満たしておらず、併せて、本件特許発明10及び11に係る特許請求の範囲の記載は、特許法第36条第6項第1号に適合しておらず、 <b>特許法第36条第6項第1号に規定する要件を満たしていないことから、特許法第123条第1項第4号の規定</b> によって無効とされるべきものである。	
基準 翻訳	Therefore, the patent specification fails to conform to the requirement of Article 36(4)(i) of the Patent Act, and the recitation of the Claims according to patent inventions 10 and 11 fails to conform to Article 36(6)(i) of the Patent Act; <b>i.e., fails to satisfy the requirement provided in Article 36(6)(i) of the Patent Act. Therefore,</b> the patents should be invalidated under the provision of <b>Article 123(1)(iv) of the Patent Act.</b>	
み自 ①	Therefore, the disclosures in the description of the present application do not satisfy the requirements specified in claim 36 and claim 4 of the present application, and the disclosures in the claims of the present application as in claims 10 and 11 of the present application do not comply with claim 36 and claim 6 of the present application, and therefore are considered to be invalid by the stipulations of <b>claims 123 and 4 of the first claim,</b> and therefore are considered to be invalid <b>by the stipulations in claims and 1 of the claim of the present application. 6 1 1 36.</b>	フレーズ単位で対訳になっている点は「み自②」よりもレベルが高いが、「第1項」が「first claim」になる等、個々の訳語が不適切。
み自 ②	Therefore, the description in the description of the present application does not satisfy the requirements defined in claim 36 and claim 4 of the present application, and the disclosures in	「特許法第123条第1項第4号」が2箇

	<p>the claims of the invention as in claims 10 and 11 of the present patent application do not comply with the invention as in claim 36 of the present application, and the requirement defined in claim 36 of the present application does not satisfy the requirement specified in claim 1 of the present application, and therefore, the invention as in <a href="#">claim 123</a> is a matter that should be invalidated <b>by the stipulations in claims, 1 and 4 of the section of the present application. 6 1 6 1.</b></p>	<p>所に分かれている等、「み自①」よりも訳質が悪い。</p>
み自③	<p>Accordingly, the description of the patent specification does not meet the requirement stipulated in Article 36 (4) (I) of the Patent Act, and the description of the scope of claims according to Patent inventions 10 and 11 does not comply with Article 36 (6) (I) of the Patent Act <b>and does not meet the requirement stipulated in Article 36 (6) (1) of the Patent Act. Therefore,</b> the patent should be invalidated under the provisions of <a href="#">Article 123 (1) (4) of the Patent Act. 1 1.</a></p>	<p>語順も個々の訳語も基準訳に近い。</p>
み自④	<p>Therefore, the description of the Description of the Patent does not meet the requirement stipulated in Article 36 (4) (I) of the Patent Act, and the description of the scope of claims relating to the patent inventions 10 and 11 does not comply with Article 36 (1) (I) of the Patent Act, <b>and does not meet the requirement stipulated in Article 36 (6) (I) of the Patent Act. Therefore,</b> the description of the Patent specification should be invalidated under the provisions of <a href="#">Article 123 (1) (4) of the Patent Act. 1 6 1.</a></p>	<p>「み自③」と同レベルで訳質の改善が見られる。</p>

学習コーパスを増やしても訳質が改善されない例

例)

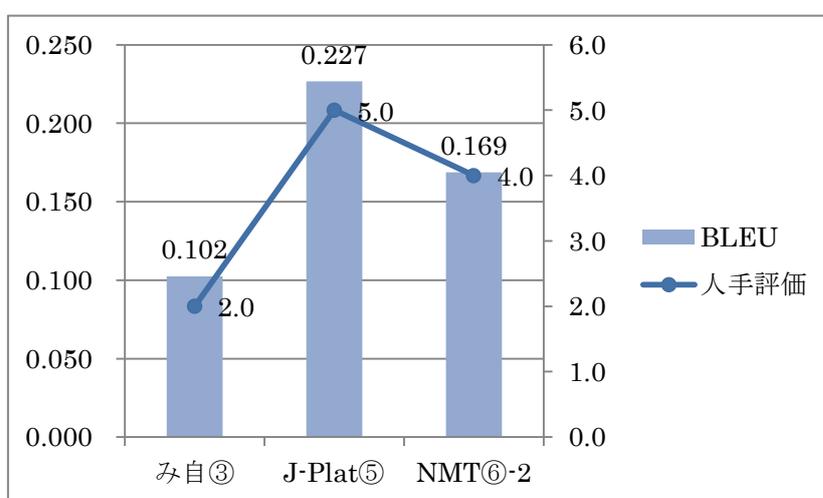


原文	しかし、前記 2(9)~(12)のポリイミドフィルムについては、 <b>当業者が</b> 、本件特許明細書の記載及び本件出願当時の技術常識に基づき、これを実施することができない。	
基準翻訳	However, <b>a person skilled in the art fails to have a specific reason</b> to manufacture a polyimide film of the aforesaid items 2(9) to (12) on the basis of the descriptions of the patent specification and the common technical knowledge as of the filing.	
み自①	However, <b>a person skilled in the art could not carry out</b> the polyimide films of 2 (9) - (12) on the basis of the disclosures of the description of the present application and common general technical knowledge at the time of the filing of the present application.	「み自①」「み自②」とともに、主語と述語は基準翻訳とほぼ同じ構成となっている。
み自②	However, <b>a person skilled in the art could not carry out</b> the aforementioned 2 (9) - (12) polyimide film on the basis of the disclosures in the description of the present application and common technical knowledge at the time of filing of the present application.	
み自③	However, <u>the polyimide films of the aforementioned 2 (9) to (12)</u> <b>cannot be carried out</b> on the basis of the descriptions in the description of the Patent and the common general technical knowledge at the time of the filing of the patent application.	「み自③」「み自④」とともに、主語である「当業者」の訳がな

み 自 ④	However, <u>the polyimide film of the aforesaid 2 (9) to (12) cannot be carried out</u> based on the description of the description of the patent specification and the common general technical knowledge at the time of the filing of the patent, on the basis of the polyimide film of the description of the Patent.	くなり、「前記2(9)~(12)のポリイミドフィルム」が主語になっている。
-------------	--	---------------------------------------

学習コーパスを増やしたことで訳質が改悪されてしまった例

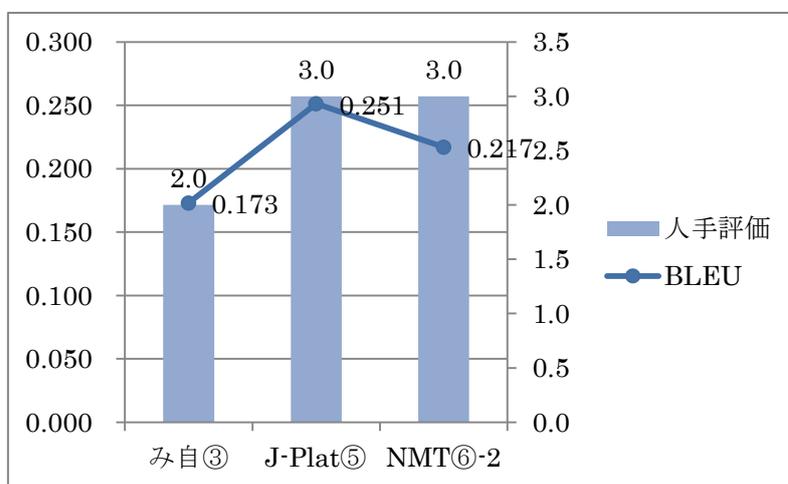
例 1)



原 文	以上によれば、「infiltrate」と「invade」が、基底膜を突き破るか否かで明確に使い分けられているとの被告の主張を認めるに足りる証拠はないというべきである。	
基 準 翻 訳	As seen above, there is no ground sufficient to accept the Defendant's argument to the effect that "infiltrate" and "invasion" are distinguished from each other with a criterion of whether or not a basement membrane is penetrated.	
み 自 ③	According to the above, it should be said that there is no evidence that it should be reasonably used for the demandee to recognize whether or not the <u>iningrate</u> is definitely used by the base membrane.	「み自③」では「infiltrate」と「invade」が訳抜けになり、原文にない「iningrate」が出現する。

J-Plat ⑤	According to the above, it should be said that there is no evidence which is sufficient for accepting the opinion of a defendant <b>that "infiltrate" and "invade" are properly used clearly by whether basement membrane is broken through.</b>	この3訳の中では最も基準訳に近い。
NMT ⑥-2	In view of the above, it should be considered that there is insufficient evidence <b>that "infiltrate" and "invade" can be clearly used in the light of whether or not it is determined whether or not a base membrane is broken.</b>	「NMT⑥-2」では「被告の主張を認めるに足りる」の対訳がない。

例2)



原文	そして、引用発明のラロキシフェンを活性成分として含む医薬製剤も、当然、備えている性質であると認められる。	
基準訳	It can be also recognized that such inherent properties of the pharmaceutical formulation can be those naturally imparted to a pharmaceutical formulation comprising raloxifene of Cited Invention as an active ingredient.	
み自③	In addition, it is found that a pharmaceutical preparation containing phenoxazine in the cited invention as an active ingredient is a property that <u>a person skilled in the art could easily achieve as a matter of course.</u>	「み自③」では原文にない「a person skilled in the art could easily achieve」の表現が出現する。

J-Plat ⑤	And naturally a pharmaceutical formulation which contains raloxifene of a Cited Invention as an active ingredient is also accepted to be the property which it has.	「J-Plat⑤」と「NMT⑥-2」はどちらも比較的基準訳に近い構成で訳されている。
NMT ⑥-2	In addition, a pharmaceutical preparation containing raloxifene in the cited invention as an active ingredient is also considered to have a property that is naturally provided.	

## 7.2 審決の日英機械翻訳文提供に向けての提言

今回の調査で、高精度な審決の日英機械翻訳文を提供するには、「み自③」や「み自④」のように、見解書と審決文の両方を含む学習データを一定量以上適用した NMT を用いることが有効であることが分かった。今後、審決の機械翻訳文の精度をさらに向上させるために有効と思われる点について以下に記載する。

### 提言 1 審決文の日本語の書き方の改善

下記に例示した日本語表現の問題点を改善することで、機械翻訳文の精度を測ることが可能と思われる。

例)

- 長文を複数の短文に分けて、各文の因果関係を明確にする（「～たり、～たり」等の繰り返し表現は避ける）。
- 主語や目的語を省略せずに明示する。
- 主題成分や長い修飾成分を明示するために、読点で区切る表現に言い換える。
- 並立表現の並立要素が同じ表現になるように整える。
- 冗長な表現を避けて簡潔な表現を用いる。

### 提言 2 学習量を増やす

本調査において、NMT では学習量が多くなるに従って機械翻訳文の精度も向上することが明らかとなったが、対象となる文書に類似した文書を多く含む素材を学習データとすることで、機械翻訳の品質をより効率的に改善できると考えられる。

NMT を用いた審決の機械翻訳文においても、従来の特許公報に加えて、見解書や審決文等の審査関連書類を数多く学習データに適用することで、機械翻訳文の精度がより向上すると思われる。

### 提言 3 学習コーパスを整備する

NMTに学習データを適用する際はデータ量を増やすだけでなく、個々の学習データの精査も必要になるものと思われる。

本事業で見解書コーパスを学習データに適用する際にも、下記の条件に合致したものを除外してNMTの機械翻訳の品質が不用意に低下しないよう配慮した。(除外したものについては「8. 付録」の「8.1 見解書コーパス」を参照されたい)。

- ・日英の対訳になっていないもの
- ・文長が1,000バイト以上のもの
- ・鉤括弧の開閉が不完全なもの(どちらか一方が欠けているもの)

対訳になっていないコーパスを学習させることで、原文と対応していない訳語が生成される等の不具合が生じることもあるので、学習データを事前に精査し、翻訳結果に悪影響を及ぼすことが予想されるものは学習データから除外することを推奨する。

### 提言 4 機械翻訳の前後処理

今回の調査で最も高い評価となったNMTを用いる場合、同じフレーズや語句が繰り返し出現する『湧き出し』が大きな課題となることが予想される。この湧き出しについては、文節が長く、繰り返し表現や記号・コンマ・読点が列挙される原文である場合に多く発生することが分かっているため、「提言 1」に記載したように、審決文の日本語の書き方を改善することで発生頻度の低下を図る等の対応が必要となる。

また、NMTでは『訳抜け』も多く発生するが、『訳抜け』についても特に原文が長文である場合に発生しやすいことが明らかとなっている。

こうした『湧き出し』や『訳抜け』への対応として、例えば、日本語原文が複文から構成される長い一文である場合に、意味上の区切り位置に改行を挿入し、機械翻訳後に、改行で分けられた部分の翻訳結果を連結するといった処理を行う方法も考えられる。

## 8. 付録

審決文 46 件を対象に実施した自動評価について、「3. 審決の機械翻訳の自動評価」に記載した「みんなの自動翻訳@TexTra®」の「み自①」～「み自④」の違いを説明する。

「表 3-1 自動評価対象の 7 種の機械翻訳」に示したように、審決文 46 件を対象に行った自動評価において 7 種の機械翻訳を用いたが、このうち、「みんなの自動翻訳@TexTra®」には下記の 4 パターンの学習データを適用した。

機械翻訳 No.	翻訳サイト又はエンジンの種類	適用した学習データ範囲
み自①	みんなの自動翻訳@TexTra®	見解書 1/2 程度
み自②	みんなの自動翻訳@TexTra®	見解書すべて
み自③	みんなの自動翻訳@TexTra®	見解書すべて、過去事業の審決全件、及び現事業の審決 37 件
み自④	みんなの自動翻訳@TexTra®	過去事業の審決全件、及び現事業の審決 37 件

表 8-1 「みんなの自動翻訳@TexTra®」に適用した学習データの概要

「み自①」～「み自④」に適用した見解書コーパスと審決コーパスの詳細について以下に説明する。

### 8.1 見解書コーパス

学習用として登録した見解書コーパスの分量は以下のとおりである。

1, 773, 809 行

貸与された見解書コーパス 2, 477, 997 行から下記を除外して 1, 773, 809 行を学習データとした。

#### 【除外した見解書コーパスの内訳】

- ・日本語文・英語文がともに同じとなる重複データは削除  
→87, 913 行を除外
- ・見解書コーパスは日本語と英語が 1 文対 1 文で対応しているもののみ使用  
→597, 082 行を除外
- ・文の対応具合を示すスコア値がゼロのデータは除外  
→6, 290 行を除外
- ・文長が 1, 000 バイト以上のもの  
→12, 903 行を除外

## 8.2 審決コーパス

学習用として登録した審決コーパスの分量は以下のとおりである。

全体：70,908 行

過去事業の審決全件、及び現事業の審決コーパスの内訳を下記表に記載する。

### [過去事業の審決全件の内訳]

審判種別	四法				対象件数	コーパス数
	特許	商標	意匠	実用		
2015 年度（平成 27 年度事業）	61	14	4	1	80	19,750
2016 年度（平成 28 年度事業）	80	15	4	1	100	20,796
2017 年度（平成 29 年度事業）	83	13	4	0	100	15,287
2018 年度（平成 30 年度事業）	29	6	2	0	37	15,075
<b>合計</b>	<b>253</b>	<b>48</b>	<b>14</b>	<b>2</b>	<b>317</b>	<b>70,908</b>

表 8.2-1 審決コーパスの学習用登録データ件数及びコーパス数

### [現事業の審決 37 件の内訳]

審判種別	四法			対象件数
	特許	商標	意匠	
無効	11	1	1	13
不服	23	2	1	26
判定	1	0	0	1
訂正	0	0	0	0
取消	0	1	0	1
異議	3	2	0	5
<b>合計</b>	<b>38</b>	<b>6</b>	<b>2</b>	<b>46</b>

表 8.2-2 審決コーパスの評価対象件数