

平成 27 年度特許庁委託事業

平成 27 年度
特許審査関連情報の日英機械翻訳文の
品質評価に関する調査
報告書

平成 28 年 3 月
一般財団法人 日本特許情報機構

目次

1. 概要.....	4
1.1 目的	4
1.2 調査内容	5
1.2.1 日英機械翻訳の翻訳精度の調査.....	5
1.2.2 日英機械翻訳文の先行技術調査における有用性調査.....	5
1.3 調査体制	5
1.4 調査スケジュール	7
1.5 調査結果概要.....	8
1.5.1 日英機械翻訳の翻訳精度の調査.....	8
1.5.2 日英機械翻訳文の先行技術調査における有用性調査.....	9
2. 日英機械翻訳の翻訳精度の調査	11
2.1 調査の実施方法	11
2.1.1 調査条件・システム環境.....	12
2.1.2 日本語原文、基準翻訳文、及び評価対象の機械翻訳文の準備	12
2.1.3 人手による評価	18
2.1.4 人手による評価結果のスコア化.....	19
2.1.5 自動評価	21
2.2 調査・分析結果	22
2.2.1 人手による評価結果の分析	22
2.2.2 自動評価結果の分析	34
2.2.3 人手による評価と自動評価の相関関係.....	42
3. 日英機械翻訳文の先行技術調査における有用性調査	48
3.1 調査の実施方法	48
3.1.1 調査条件・システム環境.....	49
3.1.2 対象文献の選定	49
3.1.3 本願の分析.....	50
3.1.4 日本語の検索式の決定	50
3.1.5 英語の検索式の決定	51
3.1.6 引用文献の機械翻訳文の準備.....	54
3.1.7 対象文献の発見可能性の検証.....	54
3.1.8 検証結果の分析	55
3.2 調査・分析結果	55
3.2.1 分野別分析.....	55
3.2.2 拒絶理由の条文別の分析.....	62

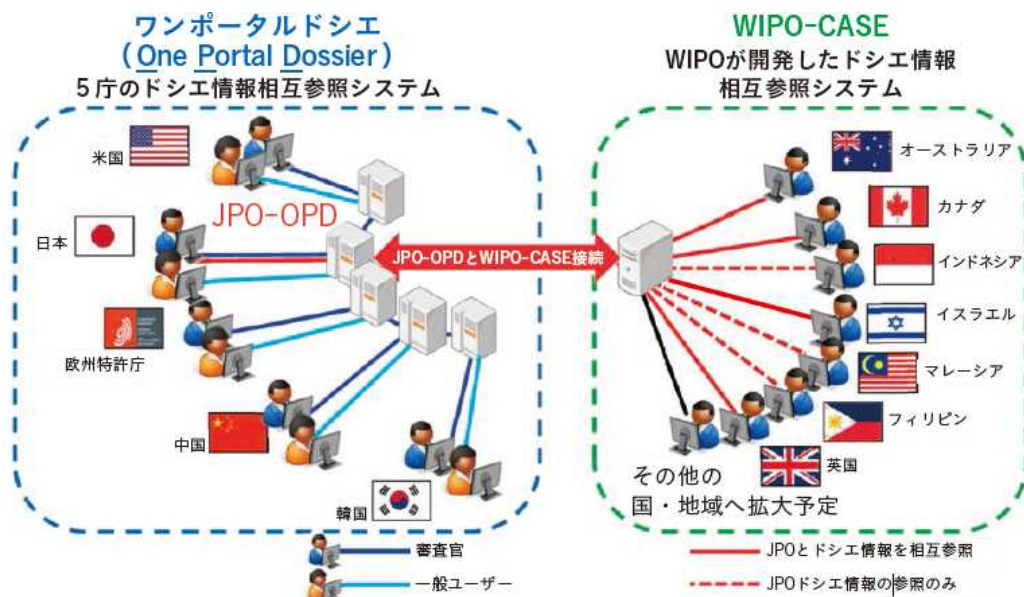
3.2.3	日英機械翻訳の翻訳精度調査結果との相関.....	62
3.2.4	PAJ の有用性分析	64
3.2.5	英語検索式でのヒット件数調査.....	67
4.	問題点・課題等.....	75
4.1	内容の伝達レベルの評価作業に関する問題点・課題.....	75
4.2	重要技術用語の翻訳精度の評価作業に関する問題点・課題.....	75
4.3	評価対象機械翻訳文の抽出における課題.....	77
4.4	英語検索に関する課題	78
4.5	拒絶理由通知書の見出し部分の翻訳不備.....	78
4.6	拒絶理由通知の引用文献の記述内容の引用	79
4.7	品質評価手順の改善案	82
4.7.1	調査対象文の機械翻訳文の取得方法.....	82
4.7.2	1 文献からの調査対象文の抽出数について	82
4.7.3	自動評価の活用	82
5.	付録.....	84
5.1	WIPO 技術分野一覧	84
5.2	自動評価手法の概要.....	85
5.2.1	BLEU	85
5.2.2	Smoothed BLEU.....	86
5.2.3	NIST.....	88
5.2.4	RIBES	90
5.3	評価者間の評価の一致度合いを調べる指標	94

1.概要

1.1 目的

特許庁では平成 16 年 10 月から「高度産業財産ネットワーク」(以下、「AIPN」という。)を通じて、我が国のサーチ及び審査結果に関する情報を英語に機械翻訳して海外特許庁に提供しており、更に、平成 25 年 7 月に運用が開始された主要 5 庁間の情報照会サービス「ワン・ポータル・ドシエ」においても、AIPN の機械翻訳機能を利用している。

これらの施策により、海外特許庁では我が国のサーチ及び審査結果を参照することで審査に係る負担が軽減されるとともに、審査の質の向上、ひいては我が国出願人の海外での円滑な権利取得につながっている。また、AIPN で機械翻訳により提供する情報は、第 1 庁で特許となった出願について第 2 庁において簡易な手続きで早期審査が受けられる施策「特許審査ハイウェイ」において活用されており、結果として海外における日本からの出願の権利取得の円滑化につながっている。



(引用) 特許行政年次報告書 2015 年版

特許審査関連情報は技術用語、専門用語が多い上に、一文が長文であることや文章の構造の複雑さから他の文章と比較して精度の高い機械翻訳が困難となっている。この問題を解決するために、特許庁では、平成 15 年度の「特許審査関連情報（ドシエ情報）の機械翻訳精度向上のための調査」によって有効な精度向上策の分析を行った。更に、平成 19 年度、平成 20 年度にそれぞれ実施した「特許審査経過情報の英語による海外特許庁への発信に係る調査」によって拒絶理由通知の記載内容の定型表現を調査して AIPN の機械翻訳システムへの辞書登録を行った。平成 22 年度に実施した「特許審査関連情報の機械翻訳による英語提供に対する精度評価に係る調査」では、辞書登録による AIPN における機械翻訳の精度向上の状況について評価を行った。

以上のような背景の下、本調査は、AIPNによる日英機械翻訳について、平成22年度以降の調査で実施された精度向上策の評価を行うとともに、他の近年進歩が著しい機械翻訳サービスによる翻訳文との精度比較や、人手により作成された公開特許公報英文抄録（以下、「PAJ」という。）と比較した先行技術調査時の有用性を評価することで、AIPNを含む海外特許庁向けの情報提供サービスにおける日英翻訳の今後のあり方について検討するための基礎資料を得ることを目的とする。

1.2 調査内容

1.2.1 日英機械翻訳の翻訳精度の調査

日英機械翻訳の翻訳精度の調査は、日本語の公開特許公報、審査書類等の日英機械翻訳文に対し、文献の種類や技術分野毎に、翻訳精度の人手による評価と自動評価をそれぞれ行うことで、日英機械翻訳の精度を分析した。

1.2.2 日英機械翻訳文の先行技術調査における有用性調査

日英機械翻訳文の先行技術調査における有用性調査は、公開特許公報の全文日英機械翻訳文が、先行技術調査における英語のキーワード検索で発見可能かどうかを具体的に検証することで、全文日英機械翻訳文の先行技術調査における有用性を評価した。

1.3 調査体制

本調査の工程の多くは専門知識を有する者が行う必要がある。例えば、本調査の「日英機械翻訳文の先行調査における有用性調査」では先行技術調査経験者の知識を基にした事業実施体制が必要となる。そこで、本調査では図1-1に示すような体制にて調査を実施した。

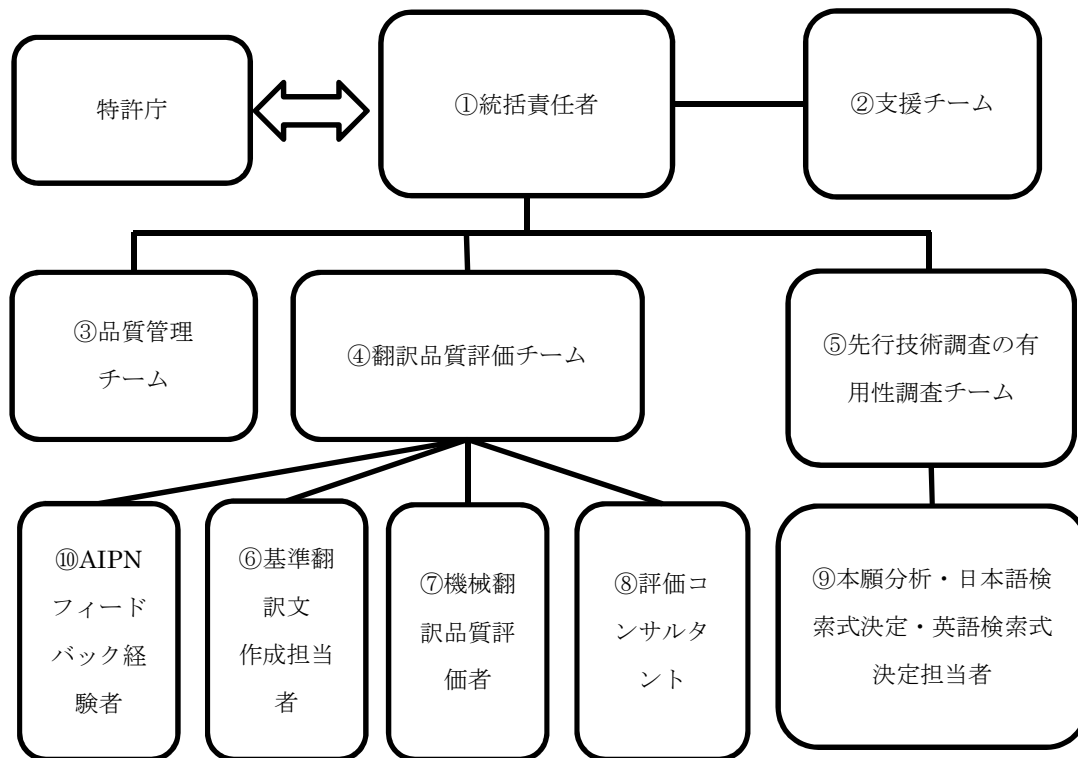


図 1-1 体制図

1.4 調査スケジュール

本調査の調査スケジュールを下図 1-2 に示す。スケジュールは大きく調査用のデータの準備作業、準備した調査用のデータの品質評価作業、及び品質評価結果の分析から構成される。これら各作業担当を分担し並行して実施可能な作業は並行して行うなど効率的に調査を実施するためのスケジュールを組んだ。

項番	項目(大)	項目(中)	項目(小)	概要	担当	2015年10月	11月	12月	2016年1月	2月	3月
1-1	データ準備	翻訳精度の調査および先行技術調査における有用性調査のデータ準備	DocDBロード	日本、米国、WPOのバックファイルをロードしてパテントファミリーの有無、米国の対応特許、技術分野(日本の筆頭IPC)の取り込み準備。	④						
1-2-1		ファミリー情報の付加	ファミリー付き本願のリストを作成	貸与された本願のリストにパテントファミリーの有無、米国とWPOの対応特許、技術分野(日本の筆頭IPC)を取り込み、ファミリー付き本願のリストを作成する。	④						
1-2-2			ファミリー付き引用文献とその出願人のリストを作成	貸与された引用文献とその出願人のリストにパテントファミリーの有無、米国とWPOの対応特許、技術分野(日本の筆頭IPC)を取り込み、ファミリー付き引用文献とその出願人のリストを作成する。	④						
2-1-1	4.1 日英機械翻訳の精度調査	①-1 公開特許公報(210文)の選定	パテントファミリーがある/ない日本の公開特許文書の選定	ファミリー付き本願のリストからパテントファミリーがある日本の公開特許文書を選定し、プリントおよび電子データを作成する。パテントファミリーがある場合、米国の公報もプリントおよび電子データを作成する。	④						
2-1-2			4.2 ①の文書取り込み	4.2 日英機械翻訳文の先行技術調査における有用性調査の文書の日本の公開特許文書のプリントおよび電子データを作成する。ファミリーがある場合、米国の公報もプリントおよび電子データを作成する。	④						
2-1-3			日本語原文の選定(630文)	上で選定した公開特許公報から、要約、特許請求の範囲及び発明の詳細な説明からそれぞれ1文ずつ調査の対象とする日本語原文を選定する。	④						
2-1-3			重要技術用語の選定	上で日本語原文を選定する際、文に含まれる重要技術用語を選定する。技術分野毎に10語選定する。	④、⑨						
2-1-4			①-2 拒絶理由通知書(70文)の選定	上で選定した公開特許公報の中から、対応する出願の審査において拒絶理由通知書が作成されている20文を選定する。	④						
2-1-5			日本語原文の選定(70文)	上で選定した拒絶理由通知書から3〜7文ずつ日本語の文を抽出する。技術分野や拒絶理由の衆文がなるべく偏らないようにする。	④						
2-1-6			重要技術用語の選定	上で日本語原文を選定する際、文に含まれる重要技術用語を追加選定する。公開特許公報由来の用語と合わせて技術分野毎に10語選定する。	④、⑨						
		①-1と①-2でマークされた箇所日本語テキストを抽出・確認			④						

項番	項目(大)	項目(中)	項目(小)	備考		2015年10月	11月	12月	2016年1月	2月	3月
2-1-7		①-3 基準翻訳文の作成	基準翻訳文(700文)	上で選定した日本語原文を英語に翻訳する。パテントファミリーがあり、対応する対訳文ある場合は、校閲の上で採用する。	⑤						
		①-3 基準翻訳文の確認		再翻訳の指示、ケアレスミスの修正。	③						
2-1-8		①-4 英語版のJ-PlatPatの機械翻訳結果の用意	機械翻訳文J(700文)	上で選定した公開特許公報の日本語原文(630文)の英語版のJ-PlatPatの機械翻訳結果を取得する。 上で選定した拒絶理由の日本語原文(70文)のJ-PlatPatの機械翻訳結果は特許庁から貸与されたデータをマージする。	④						
2-1-9		①-5 機械翻訳文1と2の用意	機械翻訳文1、2(700文)	上で選定した日本語原文(700文)を特許庁が指定する機械翻訳サービスで翻訳して機械翻訳結果を取得する。	④						
		①-4、①-5の作業の確認		誤った作業の検出。	④						
2-2-1		② 人手による評価	「内容の伝達レベル」の評価	上記機械翻訳文J、機械翻訳文1および機械翻訳文2について、「特許文書機械翻訳の品質評価手順」に従って、人手により「内容伝達レベル」の評価を行う。	⑦						
2-2-2			「重要技術用語の翻訳精度」の評価	上記重要技術用語選定で作成した重要技術用語について、「特許文書機械翻訳の品質評価手順」に従って、人手により「重要技術用語の翻訳精度」の評価を行う。	⑦						
2-3		③ 自動評価		上記機械翻訳文作成1、機械翻訳文作成2で作成した機械翻訳文について、自動評価手法3種類を用いて自動評価を行う。	④						
2-4-1		④ 翻訳精度の評価の分析	35技術分野ごとの翻訳精度の差の分析	2セットの機械翻訳文について、人手による評価と自動評価の結果を分析し、35技術分野ごとの機械翻訳精度の差を明らかにする。	④						
2-4-2			人による評価と自動評価の差の分析	人手による評価と自動評価の結果を比較し、両者に差がある場合には、差の要因と考えられる点を分析する。	④						

項番	項目(大)	項目(中)	項目(小)	備考		2015年10月	11月	12月		2016年1月	2月	3月
3-1-1	4.2 日英機械翻訳文の先行技術調査における有用性調査	①対象文献の選定	引用文献の公開特許文献70文献	ファミリー付き本願のリストからパテントファミリーがある日本の公開特許文献を選定し、プリントおよび電子データを作成する。パテントファミリーがある場合、米国の公報もプリントおよび電子データを作成する。 (イ)(ウ)の条件はファミリー付き本願のリスト、 (エ)の条件はファミリー付き引用文献とその出願人のリストにより選別する。	⑤							
3-1-2				(ア)の条件は審査官経験者が拒絶理由通知と日本と米国の公報を見て、審査官経験者が判断する。	⑤、⑨							
3-2		②本願の分析		上で選定した拒絶理由通知を分析して本願の請求項のどの発明特定事項が引用されたのかをマークする。 一人がマーク、一人が確認する。どちらも審査官経験者を配置する。	⑤、⑨							
3-3		③日本語の検索式の決定		上の本願の発明特定事項を見て①で選定した文献を検索するための日本語キーワードによる検索式を作成する。二人が独立に1式ずつ、計2式を作成する。	⑤、⑨							
3-4		④英語検索式の決定		上の日本語検索式を英訳する。	⑨							
3-5-1		⑤-1 英語版のJ-PlatPatの機械翻訳結果の用意	機械翻訳文J(70文献)	上で選定した公開特許公報の英語版のJ-PlatPatの機械翻訳結果を取得する。	⑤							
3-5-2		⑤-2機械翻訳文1と2の用意	機械翻訳文1、2(70文献)	上で選定した公開特許公報(70文献)の特許庁が指定する機械翻訳サービスで翻訳して機械翻訳結果を取得する。	⑤							
3-6		⑥対象文献の発見可能性の検証		④の英語検索式で⑤の3種類の機械翻訳文およびPAJを検索して発見可能性を調査する。	⑤、⑨							
3-7		⑦検証結果の分析		⑥の結果を4、1②及び③で得られた知見と合わせて分析して、先行技術文献としての有用性の差を明らかにする。	⑤、⑨							
4	4.3 報告書の作成				④、⑤							
5	定例打ち合わせ			毎月、計6回開催。事前資料作成1日、議事録作成1日、質問に対する資料作成3日。毎回5日X6回=1.5ヶ月								

図 1-2 調査スケジュール

1.5 調査結果概要

本調査では日英機械翻訳の翻訳精度の調査と日英機械翻訳文の先行技術調査における有用性調査の2種類の調査を行った。

日英機械翻訳の翻訳精度の調査は特許庁の機械翻訳(J-PlatPat)、Espacnet の機械翻訳(Google 翻訳)、NICT のみんなの自動翻訳(みんなの翻訳)の3種類の機械翻訳結果を人手評価と自動評価で評価した。

日英機械翻訳文の先行技術調査における有用性調査は外国の審査官が日本の引用文献を見つけることを想定して、引用文献の機械翻訳結果を英語で検索して見つけることができるかどうかを評価した。

1.5.1 日英機械翻訳の翻訳精度の調査

日英機械翻訳の翻訳精度の調査は、日本語の公開特許公報と審査書類等の日英機械翻訳文に対し、文献の種類や技術分野毎に、翻訳精度の人手による評価と自動評価を行い、日英機械翻訳の精度を分析した。

(1) 調査対象

調査対象は日本の公開特許公報 210 文献から抽出した 630 文と拒絶理由通知書 19 件から抽出した 70 文の合計 700 文に対応する J-PlatPat)、Google 翻訳、みんなの翻訳の機械翻訳文である。公開特許公報は 35 の技術分野について各分野 6 文献ずつを選定した。各文献の要約・特許請求の範囲・発明な詳細な説明から各 1 文を選定した。

また、1 分野について 10 語の技術用語をその分野の重要技術用語として抽出した。

(2)調査項目

調査対象の文を人手評価と自動評価して複数の観点で分析した。人手評価は特許庁の「特許文献機械翻訳の品質評価手順 Ver1.0」に従って、「内容の伝達レベル」と重要技術用語の評価を行った。

自動評価は BLEU、NIST 及び RIBES の 3 種類の手法で評価した。そして、3 種類の自動評価手法と人手評価の相関係数を調査した。

(3)調査結果

調査結果の要点を列記する。

- ① 3 種類の機械翻訳の内容の伝達レベルの精度は 評価が良い順に J-PlatPat > みんなの翻訳 > Google 翻訳 であった。
- ② 「内容の伝達レベル」の評価が最も良い J-PlatPat のほとんどの重要情報は正確に伝達されている割合は公報全体が 50%弱、拒絶理由が 30%弱であった。
- ③ 重要技術用語の評価結果は 3 種類の機械翻訳とも評価 A と評価 B を合わせた正解の割合が 80%程度であった。
- ④ 35 分野それぞれの評価スコアを算出して最も評価が良くない分野を特定して、原因を調査したが、原因は文の構造に起因するもので分野に特有の原因でなかった。
- ⑤ 自動評価と「内容の伝達レベル」の相関係数は低く、自動評価で人手評価を代替することはできない。

1.5.2 日英機械翻訳文の先行技術調査における有用性調査

日英機械翻訳文の先行技術調査における有用性調査は外国の審査官が日本の引用文献を見つけることを想定して、引用文献の機械翻訳結果を英語で検索して見つけることができるかどうかを評価した。

(1) 調査対象

日英機械翻訳文の先行技術調査における有用性調査での調査対象文献の選定は、日本国特許庁に対する特許出願（本願）の審査で引用された日本語の公開特許公報を J-PlatPat、Google 翻訳、及びみんなの翻訳で英語に機械翻訳したものと、本願から作成した英語の検索式を調査対象とした。

(2)調査項目

上記(1)の英語の検索式によって、引用された公開公報の機械翻訳を検索し、各機械翻訳が発見可能かの検証を行い、検証結果の分析を行った。分析は分野別、拒絶理由の条文別、及び項番 1.5.1「日英機械翻訳の翻訳精度の調査」の調査結果との関連性を調査した。また、

機械翻訳された英語を検索可能なサービス(GooglePatents 及び GPG/FX)を利用して英語の検索式による検索を行い、各検索結果の違いを調査した。

(3)調査結果

分析の結果、Google 翻訳、及びみんなの翻訳の発見可能性は、検索した検索式の 80%前後で機械翻訳を発見でき、分野間の差は少ない傾向がみられた。J-PlatPat は電気工学、化学分野の発見率が 50%前後と低く、他の分野が 60%から 70%となった。

拒絶理由条文別では、29 条 2 項が 29 条 1 項 3 号より若干高い傾向が見られた。

項番 1.5.1「日英機械翻訳の翻訳精度の調査」の調査結果との関連性を調査したが、両者の間に関連性は見られなかった。

GooglePatents 及び GPG/FX による検索結果では、GooglePatents の方が GPG/FX より多くヒットする傾向が見られた。

2. 日英機械翻訳の翻訳精度の調査

日英機械翻訳の翻訳精度の調査は、日本語の公開特許公報、審査書類等の日英機械翻訳文に対し、文献の種類や技術分野毎に、翻訳精度の人手による評価と自動評価をそれぞれ行い、日英機械翻訳の精度を分析した。

2.1 調査の実施方法

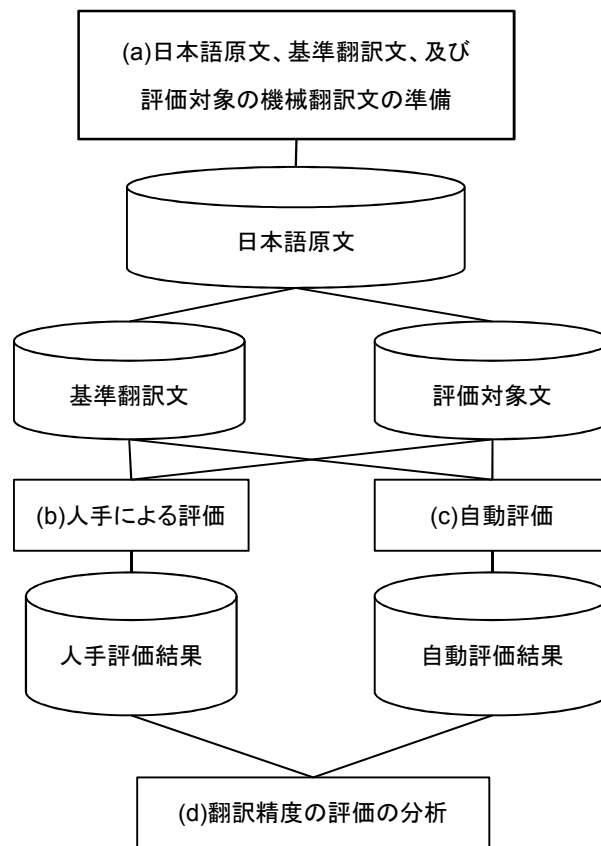


図 2-1 日英機械翻訳の翻訳精度の調査概要図

(a)日本語原文、基準翻訳文、及び評価対象の機械翻訳文の準備は、技術分野毎に公開公報と拒絶理由通知から評価対象とする日本語原文を抽出し、抽出した日本語原文を英語に翻訳して基準翻訳文を作成した。また、日本語原文に対応する機械翻訳文を抽出し評価対象文を作成した。

(b)人手による評価は、(a)で作成した基準翻訳文を参考情報として「特許文献機械翻訳の品質評価手順」に従って、評価対象文の評価を人手により行った。

(c)自動評価は、基準翻訳文を使い機械的に評価対象文の評価を行った。

(d)の翻訳精度の評価の分析は、(b)、(c)で得た評価結果についての分析を行った。

2.1.1 調査条件・システム環境

(1) 技術分野

本調査の技術分野は、WIPO が作成した「Technology Concordance Table」に基づいた 35 の技術分野(項番 5.1)を使用した。

(2) 機械翻訳文

評価対象とする機械翻訳文は、英語版 J-PlatPat、espacenet¹、及びみんなの自動翻訳²を使用し取得した。

(3) 自動評価手法

本調査で使用する自動評価手法は特許庁調査「平成 22 年度 特許審査関連情報の機械翻訳による英語提供に対する精度評価に係る調査」³で使用実績のある「BLEU」、「NIST」に加え、Workshop on Asian Translation⁴ で採用された 「RIBES」を使用した⁵。

2.1.2 日本語原文、基準翻訳文、及び評価対象の機械翻訳文の準備

調査対象となる日本語原文を抽出し、抽出した日本語について、評価の際に基準となる翻訳文の作成を行った。また、日本語原文を複数の機械翻訳システムで翻訳した。作業概要を図 2-2 に示す。作業は、

(a) 調査対象日本語原文抽出

(b) 基準翻訳文作成（翻訳）

(c) 機械翻訳文作成

のステップから構成される。

¹ espacenet から取得できない発明の名称と、要約の機械翻訳と espacenet で取得できなかった特許請求の範囲、詳細な説明の機械翻訳は google 社の機械翻訳を使用した。

² 情報通信研究機構（NICT: National Institute of Information and Communications Technology）が提供する無料のオンライン機械翻訳サービス
(<https://mt-auto-minhon-mlt.ucrjgn-x.jp/>)

³https://www.jpo.go.jp/shiryou/toushin/chousa/pdf/kikai_honyaku/h22_02.pdf

⁴ The Workshop on Asian Translation(アジア翻訳ワークショップ)はアジア言語を対象とした、評価型機械翻訳のワークショップである。(http://orchid.kuee.kyoto-u.ac.jp/WAT/)

⁵ 各自動評価手法の概要については、項番 5.2「自動評価手法の概要を参照」のこと

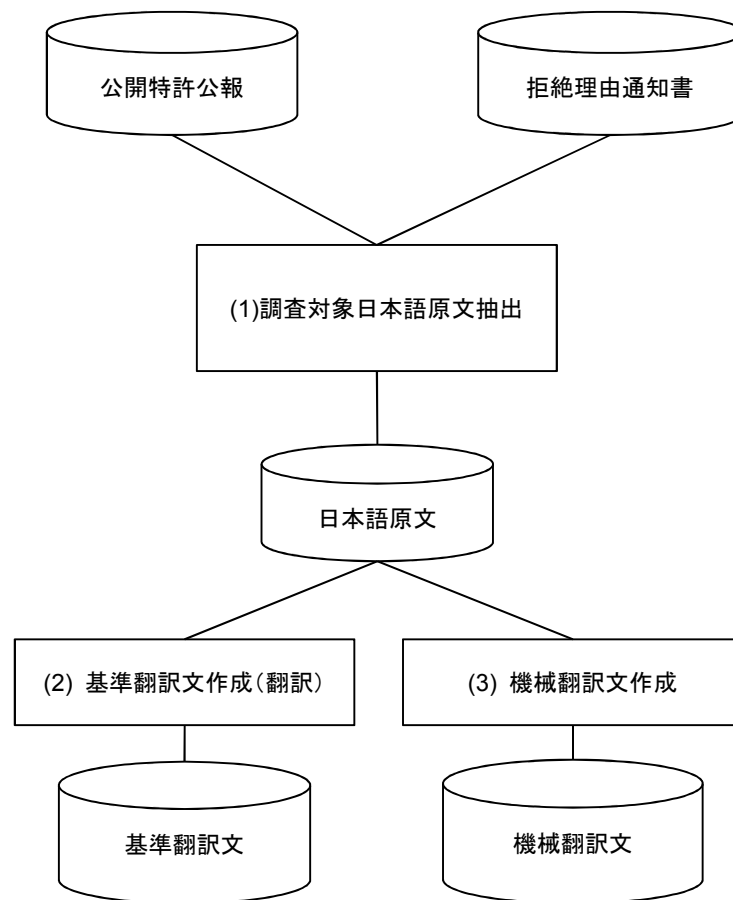


図 2-2 日本語原文、基準翻訳文、及び評価対象の機械翻訳文の準備の概要図

(1) 調査対象日本語原文抽出

評価対象となる日本語原文を、表 2-1 記載の選定方法に従い、合計 700 文を抽出した。

文献の種類	評価対象文数	評価対象文の選定方法
公開特許公報	630 文	WIPO が作成した「Technology Concordance Table」に基づいた 35 技術分野ごとに、公開特許公報を 6 文献ずつ選択し、それらの要約、特許請求の範囲、及び発明の詳細な説明の中からそれぞれ 1 文ずつ、すなわち 1 文献につき計 3 文ずつ日本語の文を抽出した。この際、各分野の 6 文献の内 2 文献は、下記表 3-1「調査対象文献選定の条件」に記載された方法により選定することとし、また、6 文献のうち半数以上はパテントファミリーの無い文献を選定した。
拒絶理由通知書	70 文	上で選定した公開特許公報の中から、対応する出

		願の審査において拒絶理由通知書が作成されている 10～20 文献を選定し、それぞれの拒絶理由通知書の中から 3～7 文ずつ日本語の文を抽出した。文の抽出には、J-PlatPat の「審査書類情報照会」を用いた。また技術分野や拒絶理由の条文がなるべく偏らないよう選定を行った。
--	--	---

表 2-1 日本語原文選定方法

文長による選定方法

調査対象の日本語文の長さは、特許庁の特許文献機械翻訳の品質評価手順 Ver.1.0 で「実際の特許文献の文長比率に従って選定することが望ましい。」とされている。そこで、本調査では上記表 2-1 の日本語原文選定方法の基準に文長による選定方法を追加した。具体的には、2014 年 1 月の公開公報と公表公報を合わせた 20,566 文献の特許明細書を母集団として文長を分析した結果に基づいて決めた。分析は要約文、特許請求の範囲及び、発明の詳細な説明毎に文長を集計し短文、中文、長文に分類した。分類は、母集団の文を文の長さ順に並べて文長が短い方から 33%を短文、文長が長い方から 33%を長文、残りを中文として、1 文の文字数範囲を取得した。項目別の短文・中文・長文の値を表 2-2 に示す。

記事	短文	中文	長文
要約文	1～55 字	56～96 字	97 字以上
特許請求の範囲	1～77 字	78～145 字	146 字以上
発明の詳細な説明	1～51 字	52～83 字	84 字以上

表 2-2 文長の分類および 1 文の文字数範囲

前記表 2-1 日本語原文選定方法では「技術分野ごとに公開特許公報を 6 文献ずつ選定し、それらの要約、特許請求の範囲、及び発明の詳細な説明の中からそれぞれ各 1 文ずつ・・・」と規定している。この規定度と表 2-2 で分類した短文・中文・長文を組み合わせ日本語原文の抽出を行った。

組み合わせは単純には 6 文献から、短文・中文・長文を各 2 文ずつ抽出するのが理想であるが、実際には理想通りの抽出が難しい場合がある。例えば、発明の詳細な説明は 1 文献の文数が多いため、他の抽出条件を合わせても合致する適切な文を見つけられる可能性が高い。一方、要約文は 1 文献に 3 文程度しかないので、他の抽出条件が適合しても、文長の条件が合う要約文が抽出できず、3 種類の記事の文が揃わないため、要約文が抽出できなかった文献を捨てて、別の文献を最初から調べる必要が生ずる。

そこで、文長の条件が似ている要約文と発明の詳細な説明を組み合わせ短文・中文・長文を抽出した。なお、特許請求の範囲は文の性質と文長の範囲が他と大きく異なるので、

このような条件の緩和は行わない。つまり、1分野6文献を合わせて見ると、特許請求の範囲は短文・中文・長文を必ず各2文抽出し、要約文は文長の判断を考えずに各2文を抽出し、発明の詳細な説明は要約文の抽出結果と合わせて、短文・中文・長文が各4文になるように調整して抽出した。

拒絶理由通知の選定方法

拒絶理由通知の選択では、下記表2-3に示すように19件の拒絶理由書を選択し、各拒絶理由通知書から3文から4文の文を選択し、合計70文選択した。機械翻訳評価では文の種類や技術分野ごとの機械翻訳精度の差を明らかにするため、技術分野や拒絶理由の条文がなるべく偏らないよう選定を行った。また、条文別では9種類の拒絶理由通知書を選択した(表2-4)。

出願番号	文数	分野	出願番号	文数	分野
2013124145	3	3	2014185095	4	15
2013223968	3	5	2014186552	4	2
2014118211	4	28	2014190200	4	12
2014128037	4	22	2014191752	4	18
2014170754	4	6	2014196289	4	16
2014172351	3	21	2014203509	4	8
2014172698	3	9	2014203524	3	31
2014172990	3	35	2014203811	3	4
2014173280	4	24	2014204888	4	17
2014184858	5	29	合計	70	19

表 2-3 選定した拒絶理由通知の分野

条文	文数
29条1項3号	18
29条2項	34
29条柱	1
29条の2	2
36条4項	3
36条6項1号	8
36条6項2号	9
39条	2
新規事項	4

表 2-4 選定した拒絶理由通知の条文

拒絶理由通知は比較的定型的な文が多い部分とそれ以外の部分から構成される。定型的な文については、「平成 15 年度の特許審査関連情報（ドシエ情報）の機械翻訳精度向上のための調査」にて分析が行われており、その分析結果は機械翻訳システムの翻訳メモリ及び辞書に反映済と考えられる。そのため、これら定型的な文を再度本調査の評価対象文とするのは適切で無いと考えられる。そこで、本調査では拒絶理由通知書のうち定型的な文でない部分、具体的には拒絶理由通知書の「記」以下の記載を中心に文を選定した（図 2-3）。

特許出願人代理人 適用条文	×× ×× 様 第 29 条第 2 項
------------------	------------------------

この出願は、次の理由によって拒絶をすべきものです。これについて意見がありましたら、この通知書の発送の日から 60 日以内に意見書を提出してください。

理 由

この出願の下記の請求項に係る発明は、その出願前に日本国内又は外国において、頒布された下記の特許公報に記載された発明又は電気通信回線を通じて公衆に利用可能となった発明に基づいて、その出願前にその発明の属する技術の分野における通常の知識を有する者が容易に発明をすることができたものであるから、特許法第 29 条第 2 項の規定により特許を受けることができない。

記 (引用文献等については引用文献等一覧参照)

- ・請求項 1
- ・引用文献等 1-2
- ・備考

引用文献 1 には、シリカを含有するポリプロピレン系樹脂が記載されており、当該組成物は、レトルト食品を包装する用途に用いる点が記載されている。引用文献 1 に記載された発明と本願請求項 1 に係る発明を対比すると、本願請求項 1 に係る発明は、表面処理が施された合成シリカを用いているが、引用文献 1 に記載された発明にはそのような記載がない点で相違する。相違点について検討するに、引用文献 2 には表面処理をしたシリカを用いることによって、分散性が改良され耐ブロッキング性が改善する点が記載されている。したがって、シリカの分散性、アンチブロッキング性を改善するために、引用文献 1 に記載された発明において、表面処理されたシリカを用いることは、当業者が容易に想到し得るものである。

引 用 文 献 等 一 覧

図 2-3 拒絶理由通知書例

拒絶理由通知の「記」以下の記載について、記述内容にはいくつかの種類がある。例えば 29 条 1 項 3 号「引用発明の認定」、「本願発明と引用発明の対比」、「その他の判断」を表す文がある。本調査ではそれぞれ以下の文数を選定した。

文の内容	文数
引用発明認定	11
対比	9
判断	18

表 2-5

(2) 基準翻訳文作成（翻訳）

機械翻訳文の評価を行うにあたり、正解とみなせる基準翻訳文が必要となる。基準翻訳文作成では、前記「(1) 調査対象日本語原文抽出」で抽出した 700 文の日本語の文を人手により英語に翻訳して基準翻訳文を作成した。基準翻訳文を効率的に行うため、PAJ やパテントファミリーの英語文献を参考にした。

(3) 機械翻訳文作成

機械翻訳文作成では、前記「(1) 調査対象日本語原文抽出」で抽出した 700 文の日本語の文に対応する機械翻訳文を J-PlatPat、Google 翻訳及び、みんなの翻訳を用いて 3 セットの機械翻訳文を作成した。

J-PlatPat データは JPlatPat サービスから取得できる部分は J-PlatPat サービスの英語版で表示されたデータを使用し、J-PlatPat サービスから直接取得できないデータ(要約及び拒絶理由通知書)は特許庁から貸与されたデータを使用した。

Google 翻訳のデータは espacenet から取得できる部分は espacenet から、取得できない要約全件と機械翻訳文が表示されなかった公開特許公報（9 文献）の請求項と明細書は google 社の翻訳サイトで機械翻訳したデータを使用した。

みんなの翻訳は NICT の機械翻訳サービス「みんなの自動翻訳@TexTra(R)」を利用した。みんなの翻訳は翻訳オプションが選べるので、要約と明細書については「JPO+NICT 特許」、請求項については「JPO+NICT 請求項」を用いて機械翻訳した。

いずれも、機械翻訳文の取得作業は、2015/11/20～2015/12/03 の間に行った。

調査対象データ毎に利用した機械翻訳を表 2-6 に示す。

2.1.4 人手による評価結果のスコア化

内容の伝達レベルの評価

「内容の伝達レベル」の評価は評価対象文に対し、下記表 2-7 に示す 5 段階評価を行った。

5：すべての重要情報が正確に伝達されている。(100%)
4：ほとんどの重要情報は正確に伝達されている。(80%～)
3：半分以上の重要情報は正確に伝達されている。(50%～)
2：いくつかの重要情報は正確に伝達されている。(20%～)
1：文意がわからない、もしくは正確に伝達されている重要情報はほとんどない。(～20%)

表 2-7 内容の伝達レベルの評価基準

(特許庁「特許文献機械翻訳の品質評価手順書 Ver.1.0」より)

本調査で評価結果をグラフ化する際は評価の文数の分布を分かりやすくするため、全体を 100%とした帯グラフで示す。一方、帯グラフは分かりやすいが結果を数値で比較する場合、特許庁評価手順 p13 においてスコア化することが求められている。ここでは本調査におけるスコア化を説明する。

「内容の伝達レベルの評価」結果のスコア化は特許庁評価手順 p13 で下記のように書かれている。

- | |
|--|
| (1) 「内容の伝達レベルの評価」結果のスコア化 |
| ①評価者ごとに評価結果の平均点を数段階の文長別に求める |
| ②各評価者の文長別平均点から、全評価者の文長別平均点を求める |
| ③全評価者の文長別平均点を、特許文献の文長比率に基づき補正し、これを「内容の伝達レベルの評価」の評価値とする |

本調査では文長別平均点の単純な平均点を「内容の伝達レベルの評価」の評価値とする。その理由は本調査で定めた短中長の定義が公開特許公報から得られる文数の分布を基に短文 33%：中文 34%：長文 33%となるように短中長文の文長を決め、かつ各分野の短中長文の文数の比率が 6 文：6 文：6 文の 18 文となるように決めたためである（項番 2.1.2 「(1) 調査対象日本語原文抽出」を参照のこと）。したがって文長による補正は不要である。

今回の「内容の伝達レベルの評価のスコア」は評価値 5 と 4 の文数の割合とする。この理由は評価値 3 の場合、半分程度の重要情報が失われているため、読みにくく技術の理解に耐えないと見なした。また、単純に評価値 5 と 4 の文数の割合としたのは全体を 100%とした帯グラフがそのままスコアを示すことになり、理解しやすくなるからである。例えば 1

分野 18 文の内、評価値 5 が 2 文、評価値 4 が 3 文あったら、その分野のスコアは
スコア＝ $(2+3) \div 18 = 27\%$ となる。

重要技術用語の翻訳精度の評価

重要技術用語の翻訳精度評価は、各技術分野 10 語ずつ重要技術用語を選択し、選択した用語を下記表 2-8 に示す評価基準で翻訳品質評価を行った。

- A (適訳語)：人手翻訳に照らし、技術的に同義かつ一般的に用いられる訳語である。
B (可訳語)：技術用語として一般的に用いられる訳語でないが、意味はおおむね正しい。
C (誤訳語)：誤訳である。
D (不訳語)：未知語、訳漏れである。

表 2-8 「重要技術用語」評価基準
(特許庁「特許文献機械翻訳の品質評価手順書 Ver.1.0」より)

「重要技術用語」の評価結果のスコア化は特許庁評価手順 p13 において下記のように書かれている。

- (2) 「重要技術用語の翻訳精度の評価」結果のスコア化
- ① 評価対象の「重要技術用語」の総数をカウントする
 - ② 各「重要技術用語」の評価結果 A から D を集計し、総数を母数として以下の数値を算出する。
 - ・ 適訳率＝評価 A (適訳) の語数÷「重要技術用語」総数
 - ・ 可訳率＝評価 B (可訳) の語数÷「重要技術用語」の語数÷総数
 - ・ 誤訳率＝評価 C (誤訳) の語数÷「重要技術用語」の語数÷総数
 - ・ 不訳率＝評価 D (不訳) の語数÷「重要技術用語」の語数÷総数
- 上記 4 つの率を用いて「重要技術用語」の翻訳精度を算出し、5 点満点の評価値に換算する。

上記の定義では、重要技術用語のスコア化に具体的な計算式を示していないが、これは、スコア換算の際は、有用性を評価する用途に応じて、各評価値に対して異なる重み付けを適宜行う必要があるためである。

今回のスコア定義では日本語から円滑に連想できる用語を高得点にするため、適訳語 A を満点の 5 点、可訳語 B を 4 点、誤訳語 C と不訳語 D は共に 0 点とした。なお、大部分の分野と機械翻訳結果において評価 C と D の頻度は合わせて 20%以下であった。また、本調査の評価値は A が 80%、B が 10%以上であった。したがって、重要技術用語のスコアはおおよそ $5 \times 80\% + 4 \times 10\% + 0 \times 10\% = 4.4$ 程度である。

2.1.5 自動評価

前記「(3) 機械翻訳文作成」で作成した 3 セットの機械翻訳文について、3 種類の自動評価手法「BLEU」、「NIST」、「RIBES」を用いて自動評価した。これら 3 種類の自動評価手法は基準翻訳文と機械翻訳結果の文の用語を比較してスコア化する。

それぞれ、連続する用語の一致の得点を高くするなど、工夫しているが、文法的な知識や用語の意味情報を使わずに評価するため、日本語と英語の様に文法や語順が大きく異なる言語間の翻訳結果の自動評価は難しい。

各自動評価手法の概要については、項番 5.2「自動評価手法の概要」を参照のこと。

2.2 調査・分析結果

2.2.1 人手による評価結果の分析

(1) 「内容の伝達レベル」の評価結果

「内容の伝達レベル」の評価は評価対象文に対し、下記表 2-7 に示す 5 段階評価を行った。

5：すべての重要情報が正確に伝達されている。(100%)
4：ほとんどの重要情報は正確に伝達されている。(80%～)
3：半分以上の重要情報は正確に伝達されている。(50%～)
2：いくつかの重要情報は正確に伝達されている。(20%～)
1：文意がわからない、もしくは正確に伝達されている重要情報はほとんどない。(～20%)

表 2-7 内容の伝達レベルの評価基準（再掲）

（特許庁「特許文献機械翻訳の品質評価手順書 Ver.1.0」より）

(a) 公開公報と拒絶理由

本調査の評価対象文は公開特許公報から取得した文と拒絶理由通知書から取得した文から成る。図 2-5 にそれぞれの「内容の伝達レベル」の評価結果を示す。下記のグラフは評価 5 の文数、評価 4 の文数、評価 3 の文数、評価 2 の文数、評価 1 の文数を下から積み上げて全体で 100%とした。

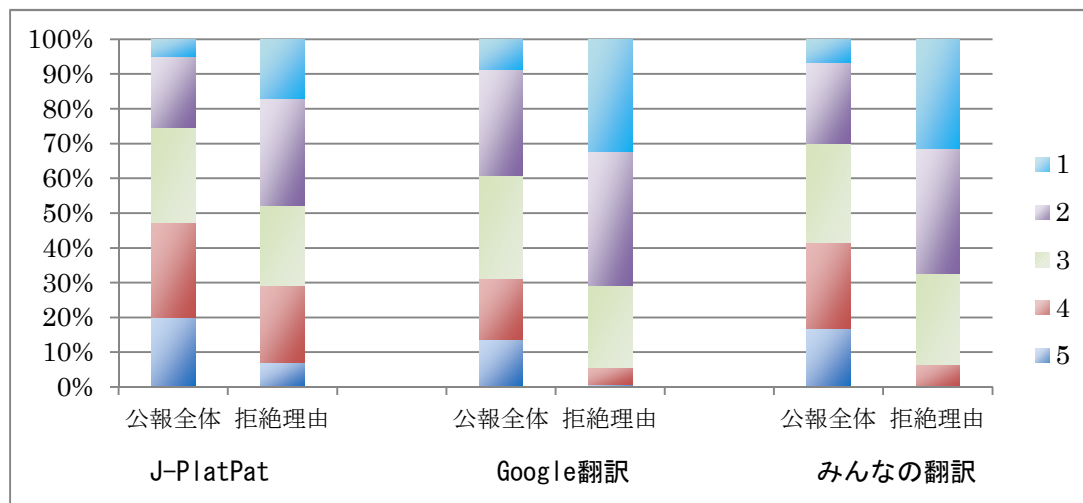


図 2-5 内容の伝達レベル（公報と拒絶理由）

上記グラフは左から J-PlatPat、Google 翻訳、みんなの翻訳を示し、機械翻訳毎に公開特許公報から取得した評価文、拒絶理由通知書から取得した評価文の評価結果となる。グラフの各帯は「内容の伝達レベル」の評価値 5 を一番下にして順に積み重ねている。例えば、J-PlatPat の公開公報の「内容の伝達レベル」5 は評価文全体のおよそ 20%である。

評価値 4 以上（ほとんど重要情報は正確に伝達されている～すべての重要情報が正確に

伝達されている)を見ると、公開公報の評価文において、J-PlatPat は 50%弱、Google 翻訳は 30%強、みんなの翻訳では 40%強となった。拒絶理由通知では評価値 4 以上の文が J-PlatPat は 20%強、Google 翻訳、みんなの翻訳ともに 10%未満となった。

以上の結果から、拒絶理由通知書の翻訳精度が公開公報の翻訳精度に比べて低いことが分かった。なお、拒絶理由通知書の翻訳精度が低い例を、項番 4.5「拒絶理由通知書の見出し部分の翻訳不備」と項番 4.6「拒絶理由通知の引用文献の記述内容の引用」に示した。

(b) 分野別 (5 セクター)

次に評価結果を技術分野別に集計した、技術分野を下記表 2-9 に示すセクター単位に集計を行い、おおまかな技術分野別の傾向を集計した。集計結果を下記図 2-6 に示す。

項番	セクター名	35 分類との対応
1	Electrical engineering (電気工学)	1～ 8
2	Instruments (機器)	9～13
3	Chemistry (化学)	14～24
4	Mechanical engineering (機械工学)	25～32
5	Other fields (その他)	33～35

表 2-9 IPC8-Technology Concordance(セクター単位)

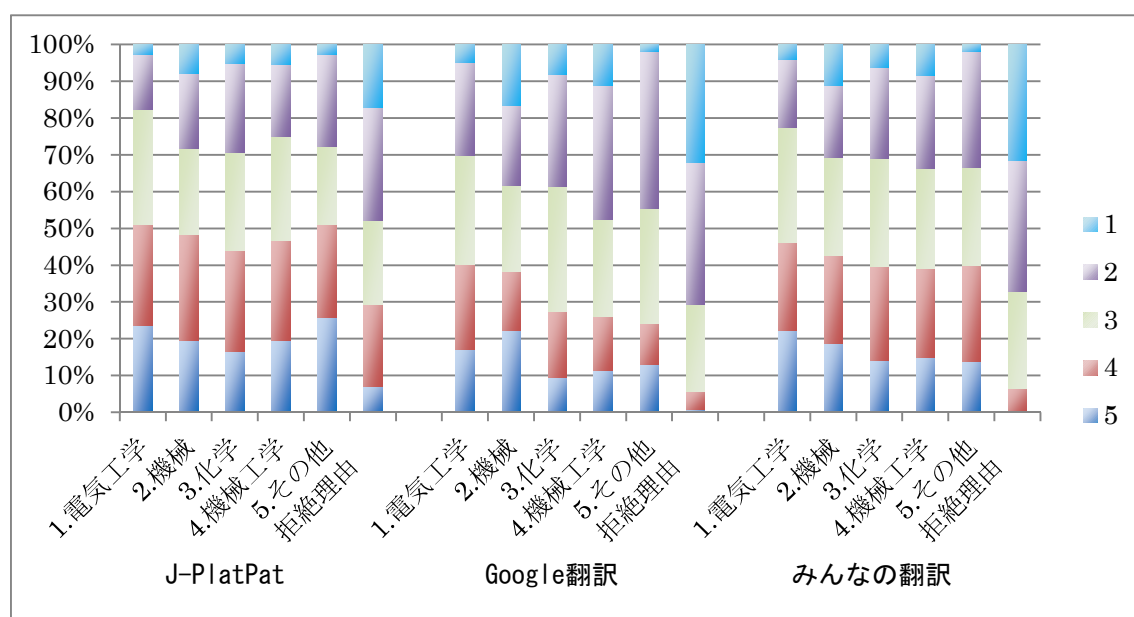


図 2-6 内容の伝達レベル (5 セクターと拒絶理由)

セクター別の評価結果を見ると、Google 翻訳の評価値が少し低目である。また、各機械翻訳とも、公報由来のセクター別の評価値に比べて、拒絶理由通知書の翻訳精度が低い。

(c) 分野別（35 分野）と拒絶理由

次に 35 の技術分野別に評価結果を集計した。集計結果を下記図 2-7 から図 2-9 に示す。

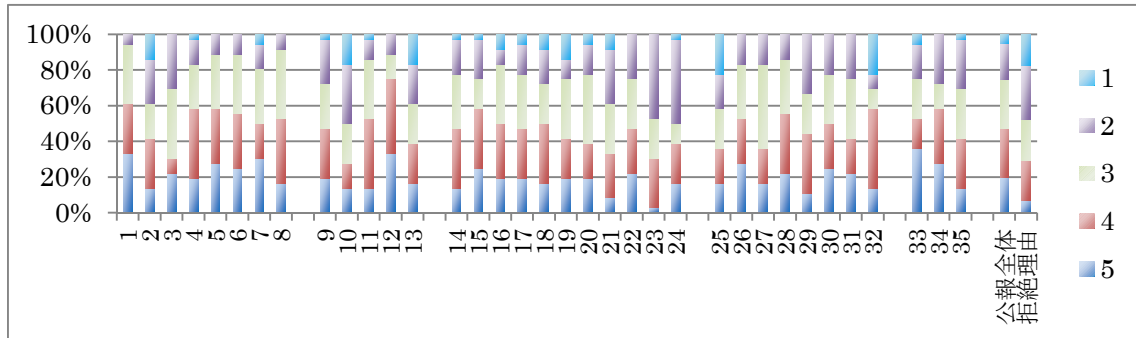


図 2-7 内容の伝達レベル (J-PlatPat)

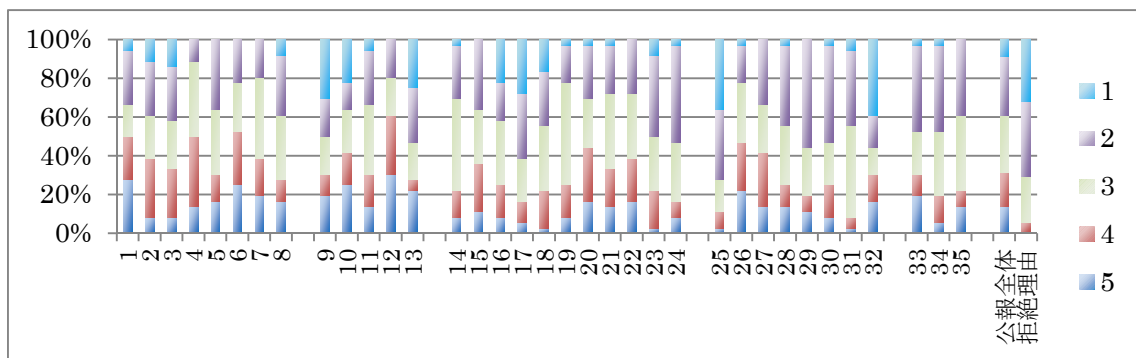


図 2-8 内容の伝達レベル(Google 翻訳)

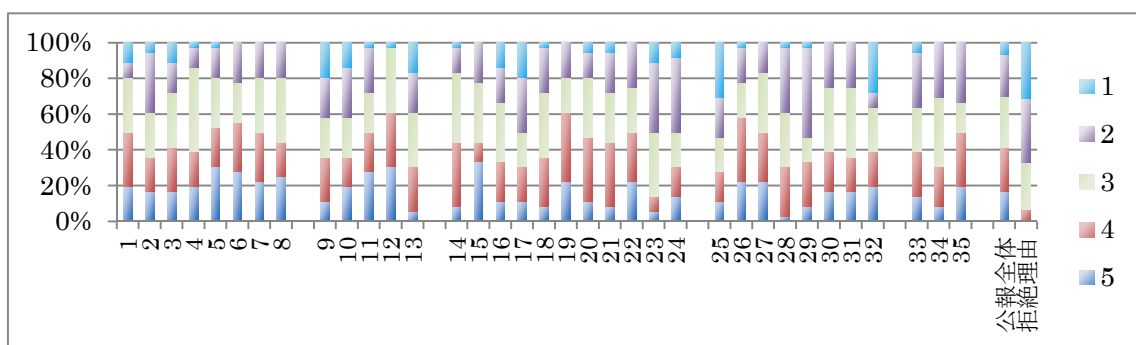


図 2-9 内容の伝達レベル(みんなの翻訳)

35 分野別の「内容の伝達レベル」は、J-PlatPat は分野 12 (control) の翻訳精度が最も高く、分野 10(Measurement)の翻訳精度が最も低くなった。Google 翻訳は分野 12(control)の翻訳精度が最も高く、分野 31(Mechanical elements)の翻訳精度が最も低くなった。みんな

なの翻訳は分野 12(control)の翻訳精度が最も高く、分野 23(Chemical engineering)の翻訳精度が低い結果となった。

これら「内容の伝達レベル」の評価が低い分野について調査対象の文のサンプルを分析した。下記表 2-10 から表 2-12 に各機械翻訳の分析例を示す。

その原因は主に原文構造の解析誤りに起因するもので、分野固有の表現（用語の翻訳品質など）に起因するためでないことが分かった。項番 2.2.1 の「(3)「内容の伝達レベル」と「重要技術用語の翻訳精度」の相関」を参照のこと。

本調査の調査結果から内容の伝達レベルの評価が低い分野についての対応策は見い出せなかった。

公開番号	JP2014241557A（分野 3）
内容の伝達レベル	2
原文	回路規模を大きくせずに、無線通信機の性能特性を、アンテナの影響を受けることなく正確に測定できる無線通信機、無線通信機の検査方法及び無線通信機の製造方法を提供する。
基準翻訳文	The present invention is to provide a radio communication device capable of accurately measuring the performance characteristic of the radio communication device without being affected by an antenna or enlarging a circuit scale, and also to provide a radio communication device inspection method and a radio communication device manufacturing method.
J-PlatPat	The inspection method of wireless radios and wireless radios which can measure the performance characteristic of wireless radios correctly without being subject to the influence of an antenna, and the manufacturing method of wireless radios are provided without enlarging a circuit scale.
備考	以下の点で、文構造の解析を誤っていたため、内容の伝達レベルは 2 になった。 ・“回路規模を大きくせず”の係先の誤り。 ・“無線通信機、無線通信機の検査方法及び無線通信機の製造方法”が並列関係にならず”及び”で並列関係になっている。

表 2-10 翻訳精度が低い例（J-PlatPat）

公開番号	JP2014240323A（分野 25）
内容の伝達レベル	1
原文	前記報知処理部は、前記乗りがご内の前記行先階ボタンを使用し、定常時の使用と異なる報知に切り替えることを特徴とする請求項 1 に記載のエレベータ装置。
基準翻訳文	What is claimed is the elevator device according to claim 1, wherein the said report process section uses the said destination floor button in the said cage, and switches to a report process different from a process in steady-state use.
Google 翻訳	The notification processing unit, an elevator apparatus according to claim 1, using the destination floor buttons in the cab, characterized in that to switch to use a different notification in the steady state.
備考	文構造分断された訳が出力されている。 （原文のカンマ単位に翻訳しそれをつなぎあわせたような訳が出力されている。）

表 2-11 翻訳精度が低い例（Google 翻訳）

公開番号	JP2014241702A（分野 1）
内容の伝達レベル	1
原文	なお、図 1（a）及び図 1（b）では、可動レールカバー 5 2 を点線で示した固定レールカバー 5 1 及び電源コード 2 と区別するため、一点鎖線で示している。
基準翻訳文	In Fig. 1 (a) and Fig. 1 (b), the movable rail cover 52 is shown by the alternate long and short dash line in order to distinguish it from the fixed rail cover 51 and the power supply cord 2 which were depicted by the dotted line.
みんなの翻訳	In FIG. 1 (a) and 1 (b), since the movable rail cover 52 to distinguish fixed rail cover 51 and power cord 2 is indicated by a dotted line and indicated by a one dot chain line.
備考	原文の構造は、”可動レールカバー 5 2 を・・・一点鎖線で示している。”となるが、みんなの翻訳ではこの構造が訳出されず、”the movable rail cover 52” ”is indicated by a dotted line and indicated by a one dot chain line”と矛盾した文意を伝えるため評価 1 となった。

表 2-12 翻訳精度が低い例（みんなの翻訳）

(d) 項目（要約、請求項、詳細な説明、拒絶理由通知）別

本調査の評価対象文は、前記表 2-1「日本語原文選定方法」にあるように、公開特許公報の「要約」、「請求項」、「詳細な説明」から取得したものと、拒絶理由通知書から取得したものとから構成されている。これら項目別に評価結果を集計したものを下記図 2-10 に示す。

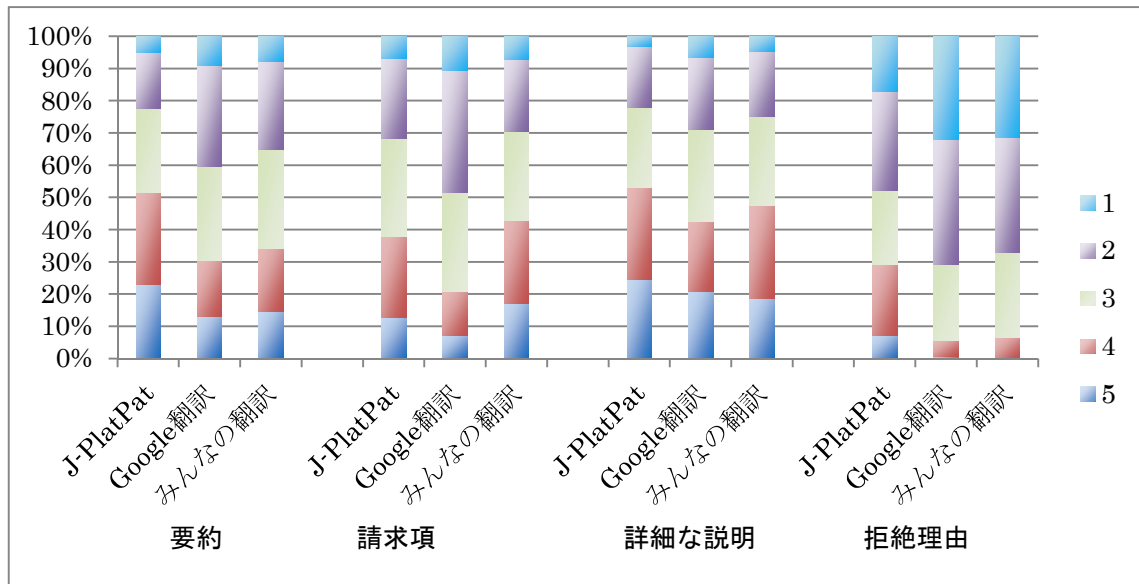


図 2-10 内容の伝達レベル・項目別

項目別に各機械翻訳の翻訳精度を評価値 4 以上の割合で比較すると、要約は J-PlatPat の精度が 50%強で他の機械翻訳の 40%強に比べ精度が高い。請求項は Google 翻訳が他の機械翻訳に比べ翻訳精度が低い。詳細な説明は、J-PlatPat の翻訳精度が高いが僅差である。拒絶理由は 3 つ全ての機械翻訳とも他の項目と比べて明らかに翻訳精度が低い。

これをルールベース翻訳と統計ベース翻訳の違いをあてはめて考えると、ルールベース翻訳は辞書と文法に基づいて学習していない文に対応できるが、統計ベース翻訳は学習していないタイプの文は翻訳精度が悪くなっていると考えられる。なお、統計ベース翻訳が学習するコーパスの量は 詳細な説明＞要約≧請求項＞拒絶理由 と考えられる。この根拠は拒絶理由の対訳コーパスが作成されていないこと、要約と請求項は国ごとに記述の傾向が異なるため、パテントファミリーから正しい対訳が採用し辛いためである。

(e) 文長別

本調査の評価対象文の文長別の翻訳精度を見る。本調査の文長は、前記表 2-2「文長の分類及び 1 文の文字数範囲」で示したように「短文」、「中文」、「長文」に分けて評価文を選定した。下記図 2-11 に文長別の評価結果を示す。

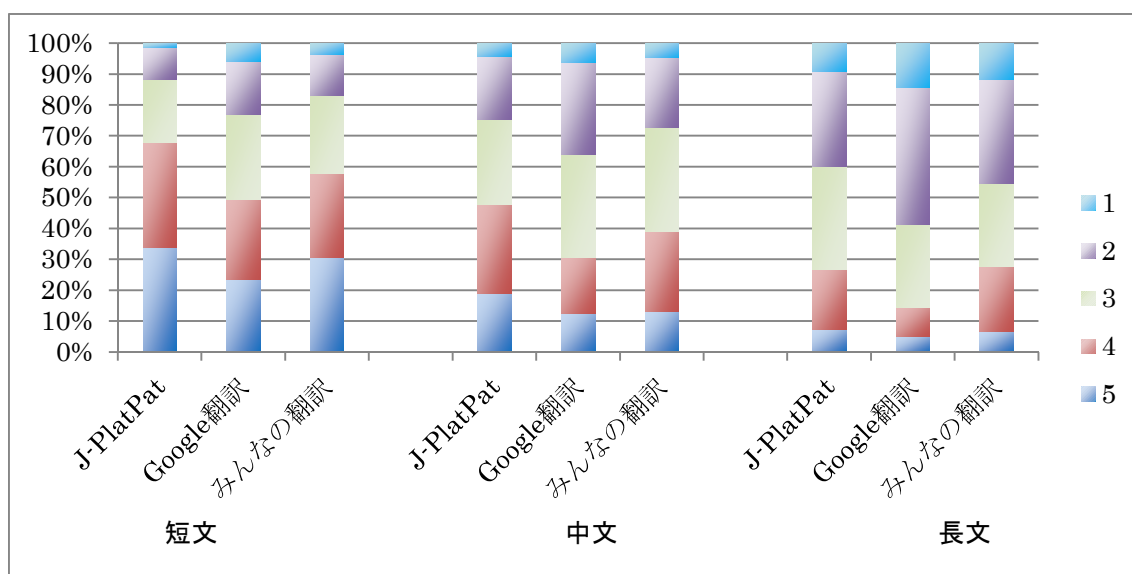


図 2-11 内容の伝達レベル・文長別

「短文」、「中文」ではJ-PlatPatの翻訳精度が一番高く、Google翻訳の翻訳精度が低い傾向が見られた。「長文」についてはみんなの翻訳の翻訳精度が一番高く、Google翻訳の翻訳精度が低い傾向が見られた。

(f) パテントファミリーの有無

パテントファミリーの有無による「内容伝達レベル」の評価結果を示す。本調査の評価対象文献(全 210 文献)は半数(105 文献)を英語のパテントファミリーを持つ文献、残りの半数は英語のパテントファミリーを持たない文献を採用した。そして、対象年はできるだけ新しい年を採用した。下記に採用した文献の公開年とその文献数を表 2-13 に示す。

	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
Fあり	0	68	0	0	2	5	4	3	1	4	3	3	4	1	1	0
Fなし	0	72	0	0	1	6	3	4	7	2	1	1	3	2	0	0
	1999	1998	1997	1996	1995	1994	1993	1992	1991							
Fあり	0	0	1	2	1	1	1	0	0							
Fなし	0	0	1	1	0	0	1	0	0							

表 2-13 評価対象文献の公開年とその文献数

下記図 2-12 にパテントファミリーの有無別の評価結果を示す。

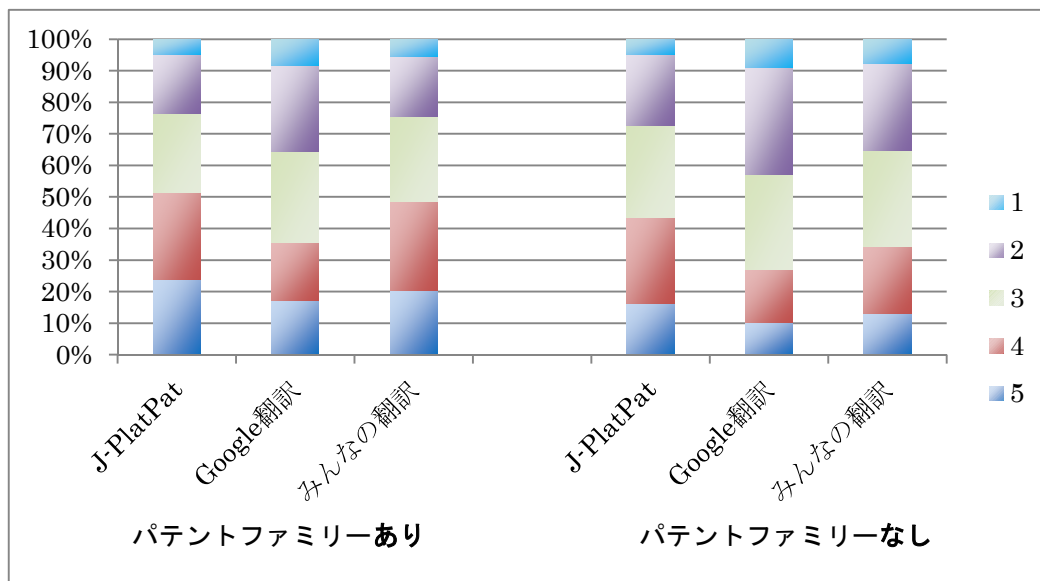


図 2-12 内容の伝達レベル（ファミリーあり・なし）

パテントファミリーありでは、J-PlatPat の翻訳精度が一番高く、みんなの翻訳がほぼ同等の翻訳精度となり、Google 翻訳の翻訳精度が低い傾向が見られた。パテントファミリーなしでは、全体のレベルは下がるが同様の結果である。

機械翻訳の方式を考えると、統計ベースの Google 翻訳とみんなの翻訳はパテントファミリーの対訳を学習するので、パテントファミリーありの方が翻訳精度が良いことは当然であるが、ルールベース翻訳の J-PlatPat もパテントファミリーありの方が翻訳精度が高い。これは翻訳に使用する辞書が対訳コーパスから採取されたと考えればパテントファミリーの有無で差があることが説明できるが、後述するように、重要技術用語の翻訳精度の差は僅差であるので、推測の域を出ない。

Google 翻訳がパテントファミリーを学習していることを確かめるため、公報の公開年毎に自動評価 BLEU の評価値の推移を見た（図 2-13、2-14）。ただし、今回の調査文献は 2014 年に集中し（表 2-10 参照）その他の年は文献数が僅かであるため、あくまで参考である。

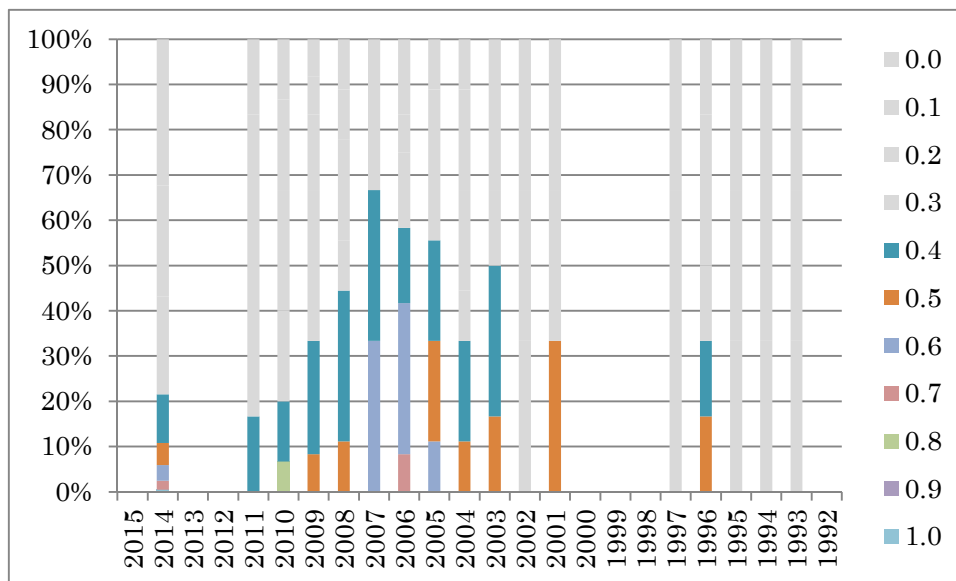


図 2-13 公開年別の Google 翻訳の BLEU 値・ファミリーあり

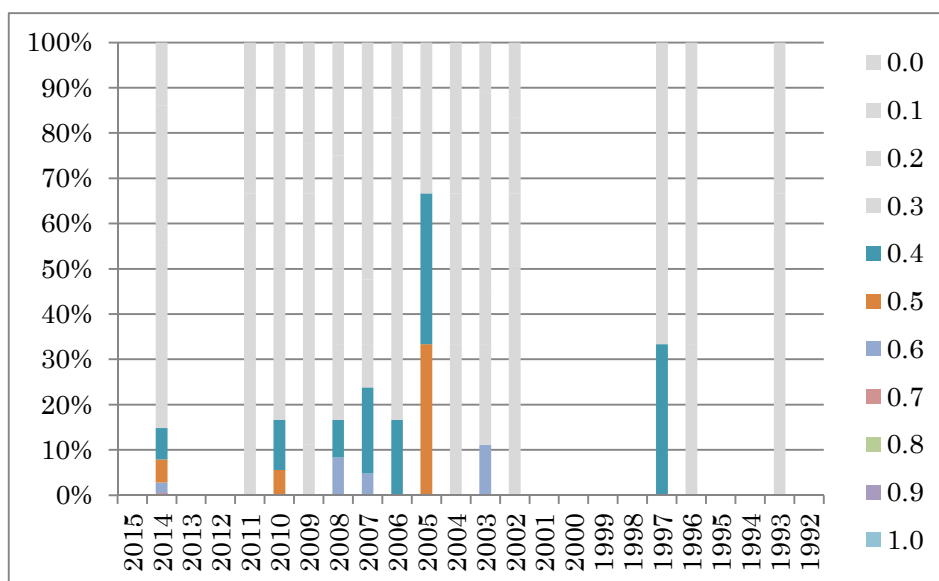


図 2-14 公開年別の Google 翻訳 の BLEU 値・ファミリーなし

この図は、BLEU 値を 0.1 単位に文数をカウントして、年ごとの構成比を帯グラフにした。今回は対訳コーパスを学習したかどうかを測るのが目的なので、BLEU の値 0.4 以上を色付けした。この観点で見ると Google 翻訳の統計ベース翻訳は 2000 年代以降を学習しているように思える。2012 年、2013 年の文が無いので確証は持てないが、2014 年は学習が追い付いていないように思える。

(2) 「重要技術用語の翻訳精度」の評価結果

重要技術用語の翻訳精度評価は、各技術分野 10 語ずつ重要技術用語を選択し、選択した用語を下記表 2-8 に示す評価基準で翻訳品質評価を行った。全体の評価結果を下記図 2-15 に、パテントファミリー有無別の評価結果を下記図 2-16 に示す。

- A (適訳語) : 人手翻訳に照らし、技術的に同義かつ一般的に用いられる訳語である。
 B (可訳語) : 技術用語として一般的に用いられる訳語でないが、意味はおおむね正しい。
 C (誤訳語) : 誤訳である。
 D (不訳語) : 未知語、訳漏れである。

表 2-8 「重要技術用語」評価基準 (再掲)

(特許庁「特許文献機械翻訳の品質評価手順書 Ver.1.0」より)

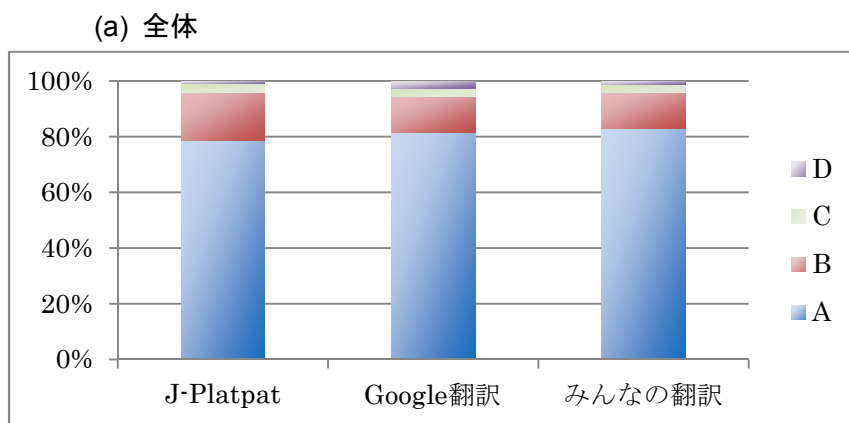


図 2-15 重要技術用語の評価結果

評価結果を見ると、評価 A (適訳語) と B (可訳語) の合計が全ての機械翻訳において 90%以上となり、機械翻訳間に大きな差は見られなかった。

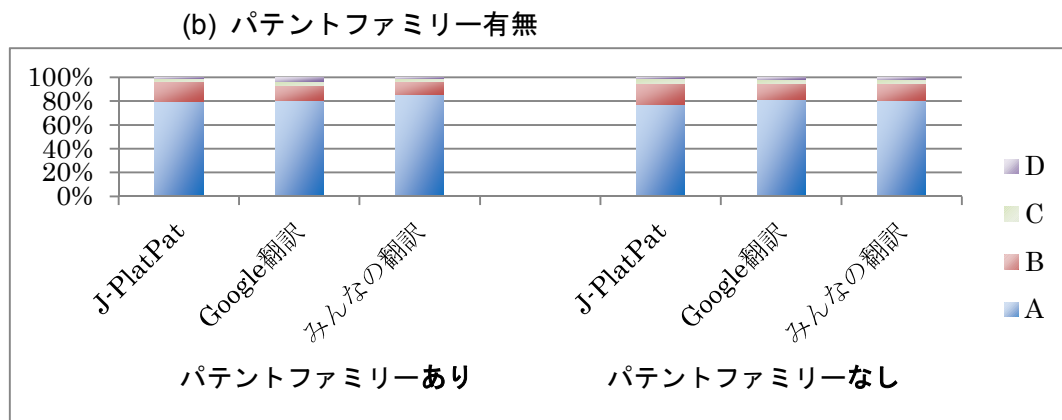


図 2-16 パテントファミリー有無・重要技術用語の評価結果

図 2-16 の重要技術用語の評価結果を見ると、3 つの機械翻訳ともパテントファミリーありの方が若干良い結果となっているが、僅差であり、この重要技術用語の評価結果からパテントファミリーありの方が「内容の伝達レベル」の人手評価値が良いことを説明することは難しい。

(c) 分野別

次に技術分野別の重要技術用語の翻訳精度を表 2-14 に示す。なお、縦軸の重要技術用語のスコアについては項番 2.1.4 「人手による評価結果のスコア化」を参照のこと。

	1	2	3	4	5	6	7	8		9	10	11	12	13
J-PlatPat	4.5	4.7	4.4	4.9	4.3	4.6	4.5	4.3		4.7	4.6	4.8	3.8	4.6
Espace	4.7	4.3	4.6	5.0	4.4	4.7	4.7	3.9		4.5	4.9	4.8	4.1	4.6
NICT	4.6	4.7	4.8	4.4	4.4	5.0	4.9	4.5		4.6	4.7	4.6	4.5	4.1
	14	15	16	17	18	19	20	21	22	23	24			
J-PlatPat	4.1	4.5	4.5	5.0	4.8	5.0	4.8	4.9	4.5	4.6	4.6			
Espace	4.9	4.5	4.5	5.0	4.4	5.0	5.0	5.0	4.9	4.9	5.0			
NICT	4.8	4.8	5.0	5.0	4.3	5.0	4.9	4.9	4.9	4.6	4.8			
	25	26	27	28	29	30	31	32		33	34	35		
J-PlatPat	4.3	5.0	5.0	4.9	4.6	4.9	4.6	4.8		4.5	5.0	4.8		
Espace	4.4	4.9	4.4	4.6	4.9	4.9	4.4	3.2		5.0	4.0	4.6		
NICT	4.3	4.6	4.9	4.4	5.0	4.9	4.9	5.0		4.9	4.0	4.3		

表 2-14 重要技術用語（分野別）

表 2-14 は 1 から 35 分野を横にならべ、機械翻訳の種類を縦にした表である。横に長いのでセクター単位が分かるように改行している。色付けしたセルは各機械翻訳のスコア下位 10%を示す。

J-PlatPat は分野 12、14、25 のスコアが低いことを示すが、最低スコアは分野 12 の 3.8 である。Google 翻訳は分野 8、32、34 のスコアが低く最低スコアは分野 32 の 3.2 である。表 2-15 「重要技術用語の評価例 Google 翻訳の分野 32(Transport)」の評価結果を見ると、「パンク」と「荷棚」の対訳に問題があり、「高速走行」は「走行」の翻訳が欠落している。Google 翻訳以外の重要技術用語のスコアは 4.8 と 5.0 なので、分野 32(Transport)の技術用語の翻訳に問題があるのは Google 翻訳だけである。一方、同じ統計ベース翻訳のみんなの翻訳は 13、18、25、34、35 のスコアが低いが最低でも 4.0 なので極端に低い分野はない。

技術用語	基準翻訳文	Google 翻訳	評価者A	評価者 B
高速走行	high-speed travel	high speed	D	D
パンク	blowout	punk	C	D
脆弱部	fragile portion	weak portion	B	C
排出量	amount of emission	emission	B	A
計器	instrument	meter	A	A
コイル抵抗	coil resistance	coil resistance	A	A
インホイールモータ駆動装置	in-wheel motor driving device	in-wheel motor drive unit	A	A
繊維強化複合材料	fiber-reinforced composite materials	fiber-reinforced composite materials	A	A
鉄道車両	railway vehicle	railway vehicle	A	A
荷棚	baggage shelf	Nitanaload shelf	C	D

表 2-15 重要技術用語の評価例 Google 翻訳の分野 32(Transport)

(3) 「内容の伝達レベル」と「重要技術用語の翻訳精度」の相関

技術分野毎の「内容の伝達レベル」と「重要技術用語の翻訳精度」の相関関係を調べ、表 2-16 を得た。得られた相関係数は絶対値 0 に近く、正負両方あることから、相関は認められない。相関が認められない原因は、文の翻訳の難しさが技術分野より調査対象文の複雑さに依ることと「重要技術用語の翻訳精度」が一つの分野で 10 語では語数が足りないためと考えられる。それぞれのスコアは項番 2.1.4 「人手による評価結果のスコア化」を参照のこと。

相関の調査対象	Pearson の相関係数
J-PlatPat の内容の伝達レベルと重要技術用語の翻訳精度の相関	-0.09
Google 翻訳の内容の伝達レベルと重要技術用語の翻訳精度の相関	0.17
みんなの翻訳の内容の伝達レベルと重要技術用語の翻訳精度の相関	0.14

表 2-16 「内容の伝達レベル」と「重要技術用語の翻訳精度」の相関

2.2.2 自動評価結果の分析

項番 2.2.1「人手による評価結果の分析」と同様に自動評価による評価結果の分析を行った。分析は 3 種類の自動評価手法毎に、分野全体、分野別、項目別、文長別、パテントファミリーの有無別の分析を行った。

(1) 公開公報、拒絶理由通知書

評価対象の公開公報、拒絶理由通知書について、自動評価を行った結果を下記図 2-17 から図 2-20 に示す。

各グラフは縦軸に自動評価値、横軸に各機械翻訳システムを取った。自動評価手法の BLEU と RIBES の評価値は 0.0 から 1.0 までの値となり 1.0 に近づくほど評価値が良いことを意味する。NIST は 0.0 以上の実数で 1.0 を超えて大きくなり、評価値の値が大きいほど評価値が良いことを意味する。

自動評価手法 BLEU、NIST では Google 翻訳の翻訳精度が高く、J-PlatPat が低い傾向が見られた。一方、RIBES ではみんなの翻訳の翻訳精度が高く、Google 翻訳の翻訳精度が低い傾向が見られた。

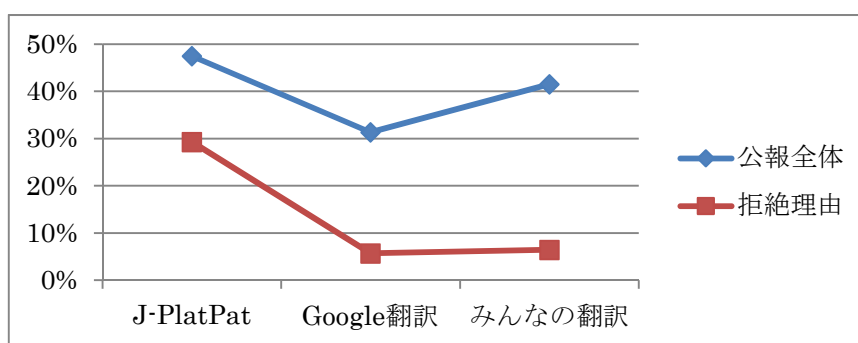


図 2-17 内容の伝達レベル（公開公報、拒絶理由通知書）

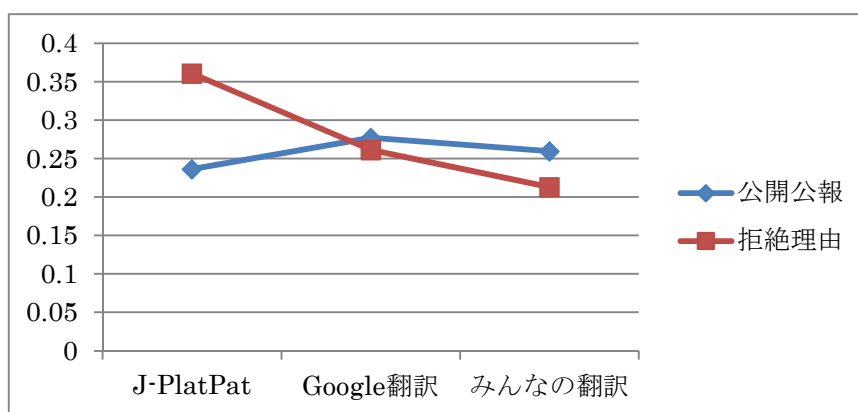


図 2-18 自動評価（公開公報、拒絶理由通知書）(BLEU)

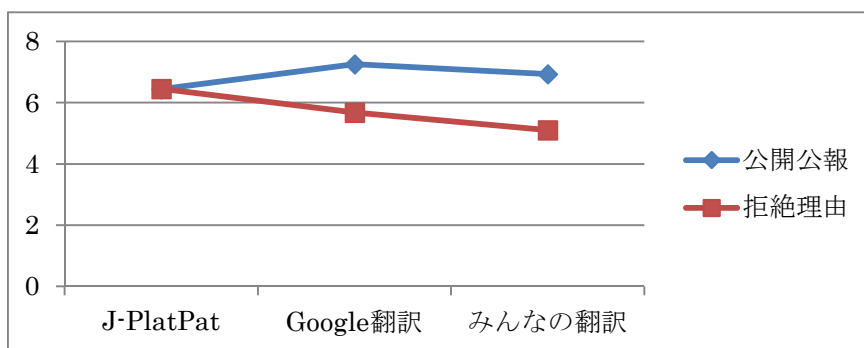


図 2-19 自動評価（公開公報、拒絶理由通知書）(NIST)

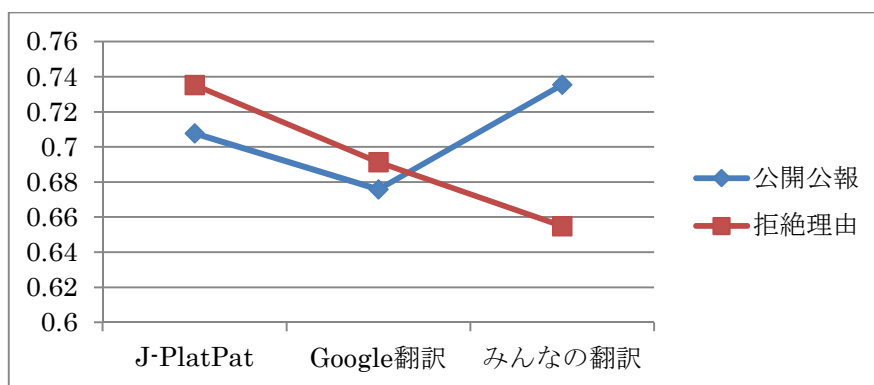


図 2-20 自動評価（公開公報、拒絶理由通知書）(RIBES)

(2) 項目（要約、請求項、詳細な説明、拒絶理由通知）別

自動評価結果を項目別に集計した結果を下記図 2-21 から図 2-24 に示す。BLEU、NIST では前記人手評価の結果と同様に詳細な説明の翻訳精度が高く、要約の翻訳精度が低い傾向が見られた。RIBES では Google 翻訳の請求項の翻訳精度が他に比べ翻訳精度が低い傾向が見られた。拒絶理由は全ての自動評価値で J-PlatPat が優れている。これは拒絶理由が他の項目と違う傾向の文であることを示している。

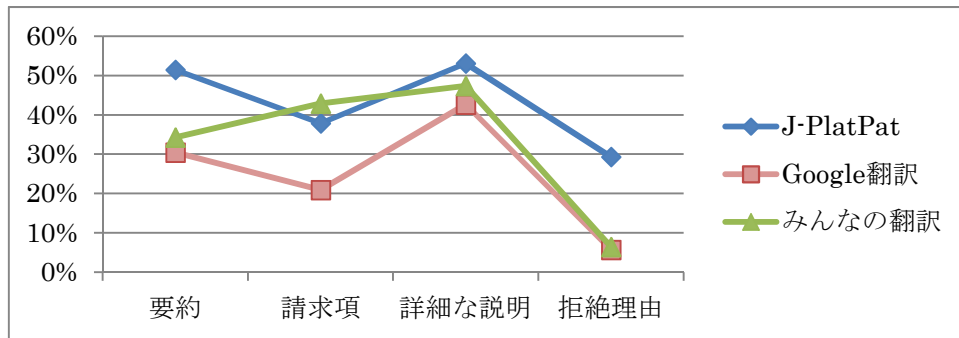


図 2-21 内容の伝達レベル (項目別)

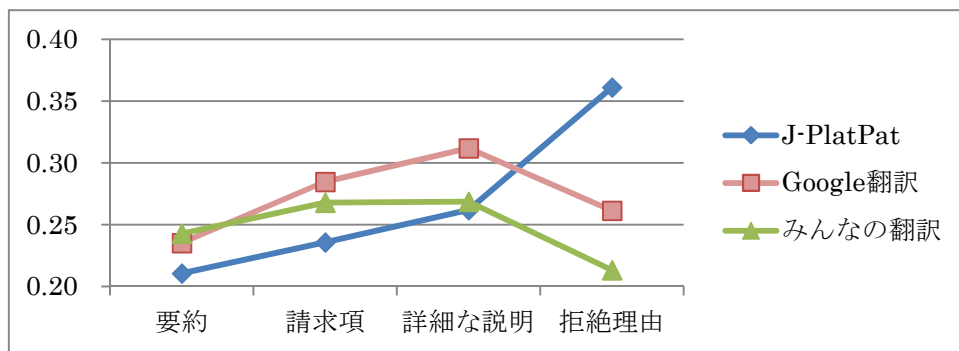


図 2-22 自動評価 (項目別) (BLEU)

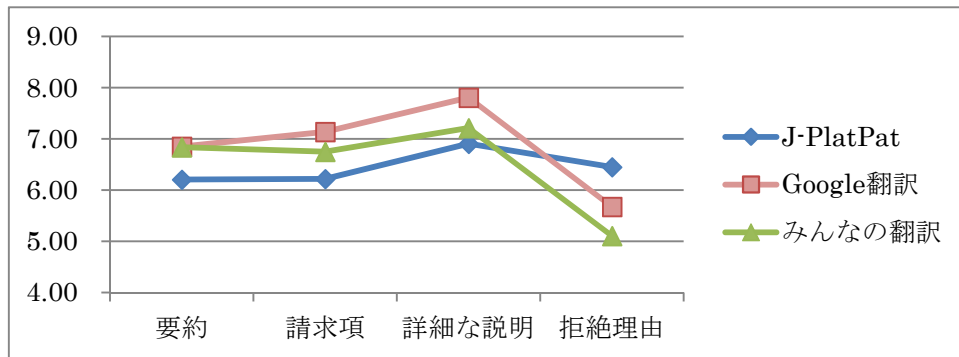


図 2-23 自動評価 (項目別) (NIST)

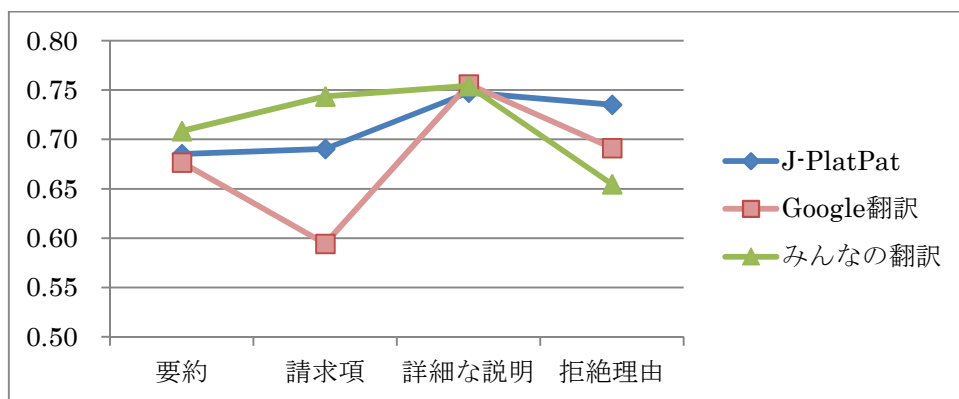


図 2-24 自動評価 (項目別) (RIBES)

(3) 文長別

自動評価結果を評価文の文長（短文、中文、長文）に分けそれぞれの評価結果を下記図 2-25 から図 2-28 に示す。BLEU、NIST では短文、中文の翻訳精度がほぼ同等となり、長文の翻訳精度が若干低くなる傾向が見られた。一方 RIBES では短文から長文になるに従い翻訳精度が低くなる傾向が見られた。BLEU、NIST の結果は項番 2.2.1 の(1)「(e) 文長別」の傾向とは異なる。これは BLEU、NIST が人手評価と明らかに異なる評価であることを示している。

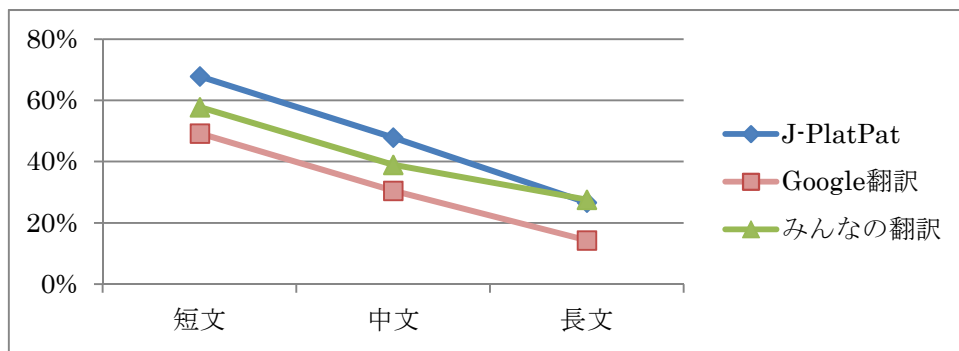


図 2-25 内容の伝達レベル (文長別)

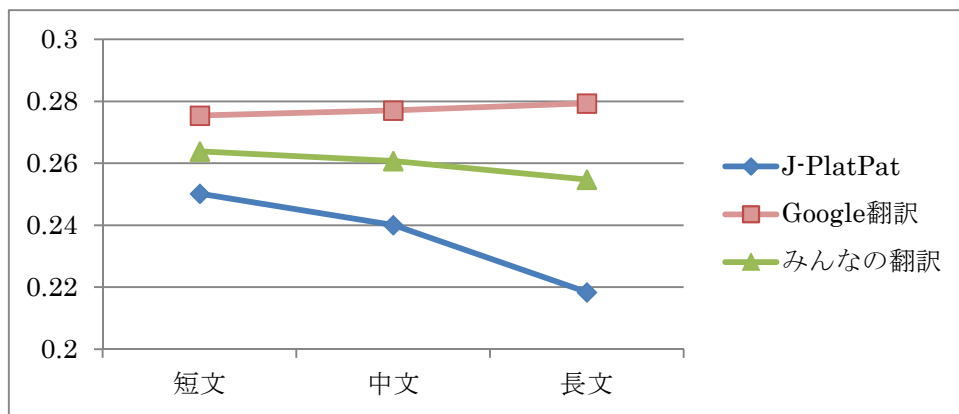


図 2-26 自動評価 (文長別) (BLEU)

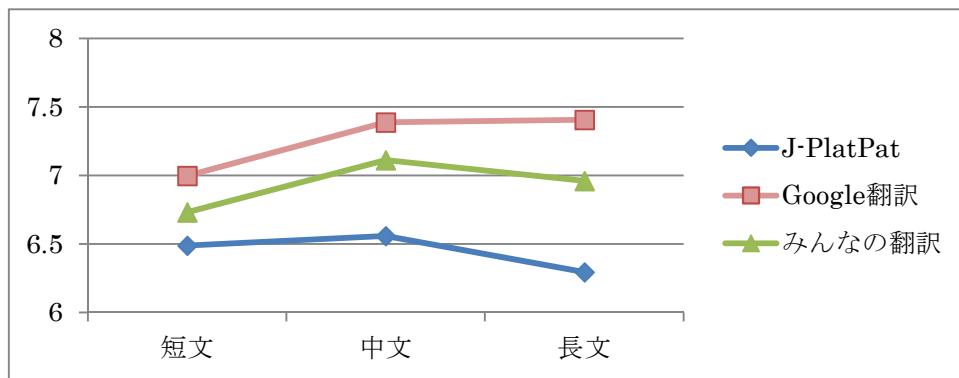


図 2-27 自動評価 (文長別) (NIST)

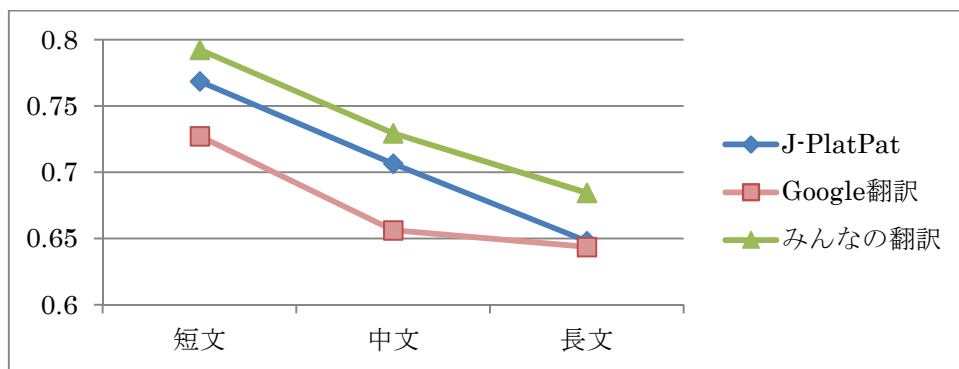


図 2-28 自動評価 (文長別) (RIBES)

(4) パテントファミリーの有無

自動評価結果を評価文の取得元文献のパテントファミリー有無に分け評価結果を集計した。集計結果を下記図 2-29 から図 2-32 に示す。統計ベース翻訳の Google 翻訳とみんなの翻訳はパテントファミリーの対訳を学習しているため、ファミリーありの方が評価値が高いのは当然である。一方、ルールベース翻訳の J-PlatPat も同じ傾向となっているが、理由は不明である。

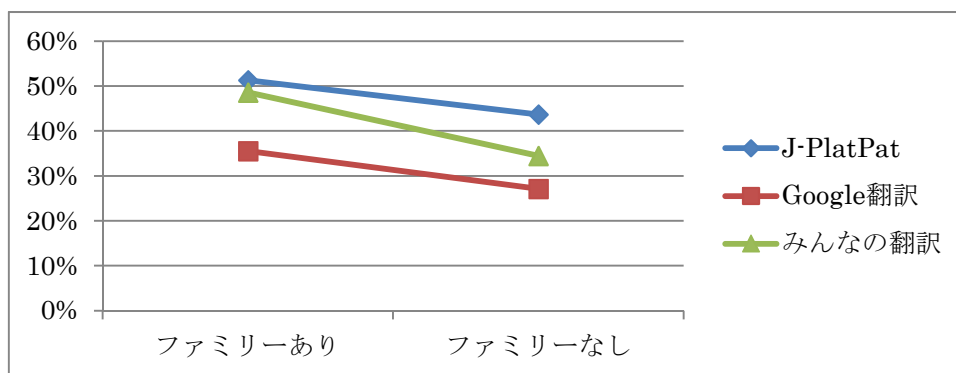


図 2-29 内容の伝達レベル (パテントファミリー有無別)

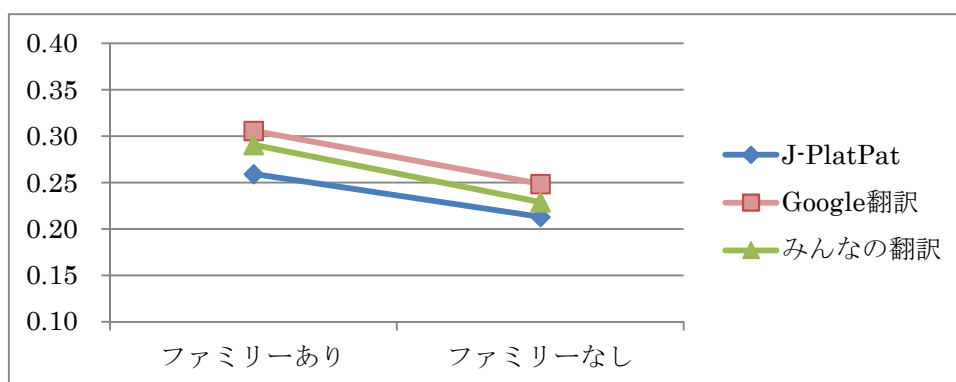


図 2-30 自動評価 (パテントファミリー有無別) (BLEU)

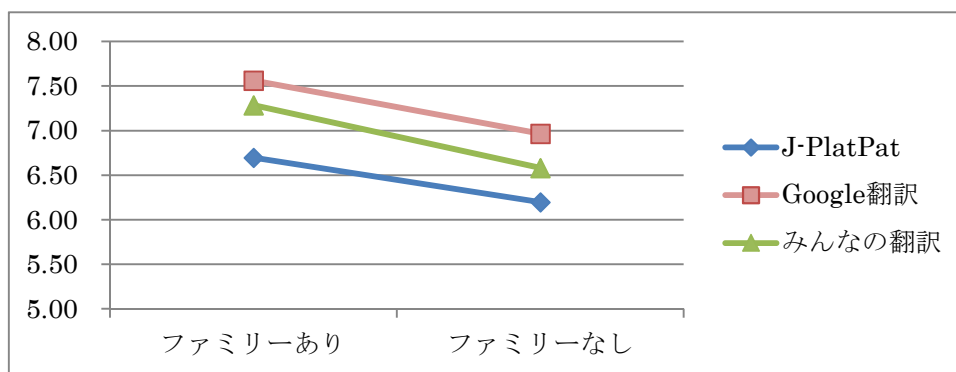


図 2-31 自動評価 (パテントファミリー有無別) (NIST)

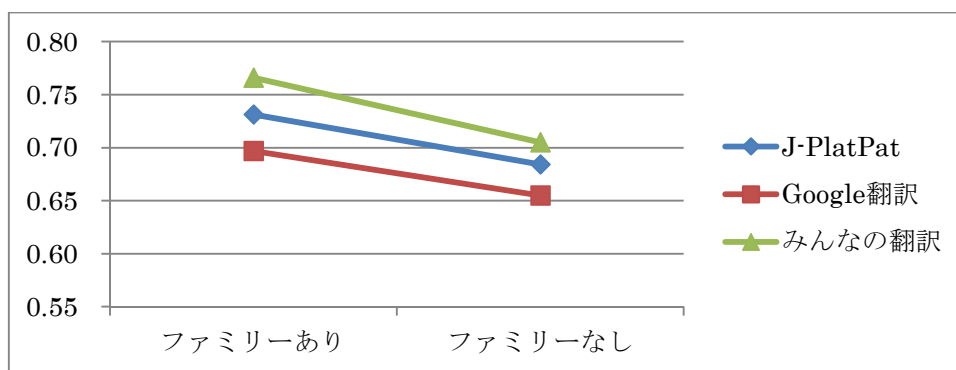


図 2-32 自動評価 (パテントファミリー有無別) (RIBES)

2.2.3 人手による評価と自動評価の相関関係

(1) 相関係数の算出

ここまでで得られた人手による文の評価結果（「内容の伝達レベル」）と文の自動評価結果（BLEU、NIST、RIBES）の相関を調べた。この相関が高ければ、コストと時間がかかる人手による評価の代わりに自動評価を使うことができる。なお、相関係数には他の機械翻訳の研究で人手による評価結果と自動評価結果の相関を見るのに用いられる Pearson の相関係数を採用した。相関を調べた評価対象文は公報 630 文の 3 種類の機械翻訳結果 1890 文の 3 種類の自動評価値（BLEU、NIST、RIBES）とその「内容の伝達レベル」の人手評価のデータである。人手評価データは評価者 2 名の評価値の平均値を利用した（表 2-17）。

相関を調べた組み合わせ	Pearson の 相関係数
3 種類の機械翻訳の BLEU と人手評価の平均（内容の伝達レベル）	0.26
3 種類の機械翻訳の NIST と人手評価の平均（内容の伝達レベル）	0.22
3 種類の機械翻訳の RIBES と人手評価の平均（内容の伝達レベル）	0.41

表 2-17 人手による評価と自動評価の相関関係

Pearson の相関係数は完全に相関する場合は 1.0 に近づき、完全に逆相関する場合は -1.0 に近づき、相関が無い場合は 0.0 に近づく。この相関係数を直感的に示すため、機械評価値を横軸にして 0.1 刻み、人手評価値 5～1 を H5～H1 の縦軸としてマスを作り、その条件に合う文数を丸の大きさで表したバブルチャートを作成した。なお、人手評価は二人の評価者の評価値の平均を四捨五入して H1～H5 とした。

バブルチャートは BLEU、NIST、RIBES の 3 種類の自動評価手法毎に作成し、図 2-33 から図 2-35 に示す。

(a) J-PlatPat の BLEU と人手評価（「内容の伝達レベル」）

J-PlatPat の BLEU と人手評価（「内容の伝達レベル」）の Pearson の相関係数は 0.26 である。BLEU の値は 0.0 から 1.0 となり、1.0 に近いほど正解文の語の組み合わせが近いことを示す。機械翻訳文の BLEU の評価値 0.0～1.0 と人手評価 H1～H5 の文数を下記のよう表にして、この表からバブルチャートを作成した。

BLEU 値が 0.2、0.3 は H1～H5 に幅広く対応するので、BLEU 値で人手評価値の高い文を効率よく抽出することはできない。

	B0	B0.1	B0.2	B0.3	B0.4	B0.5	B0.6	B0.7	B0.8	B0.9	B1.0
H1	3	3	2	0	0	0	0	0	0	0	0
H2	25	25	23	8	2	2	1	0	0	0	0
H3	47	53	53	26	14	1	3	1	0	0	0
H4	31	52	54	31	15	5	1	1	0	0	0
H5	14	33	39	24	17	11	4	4	1	1	0

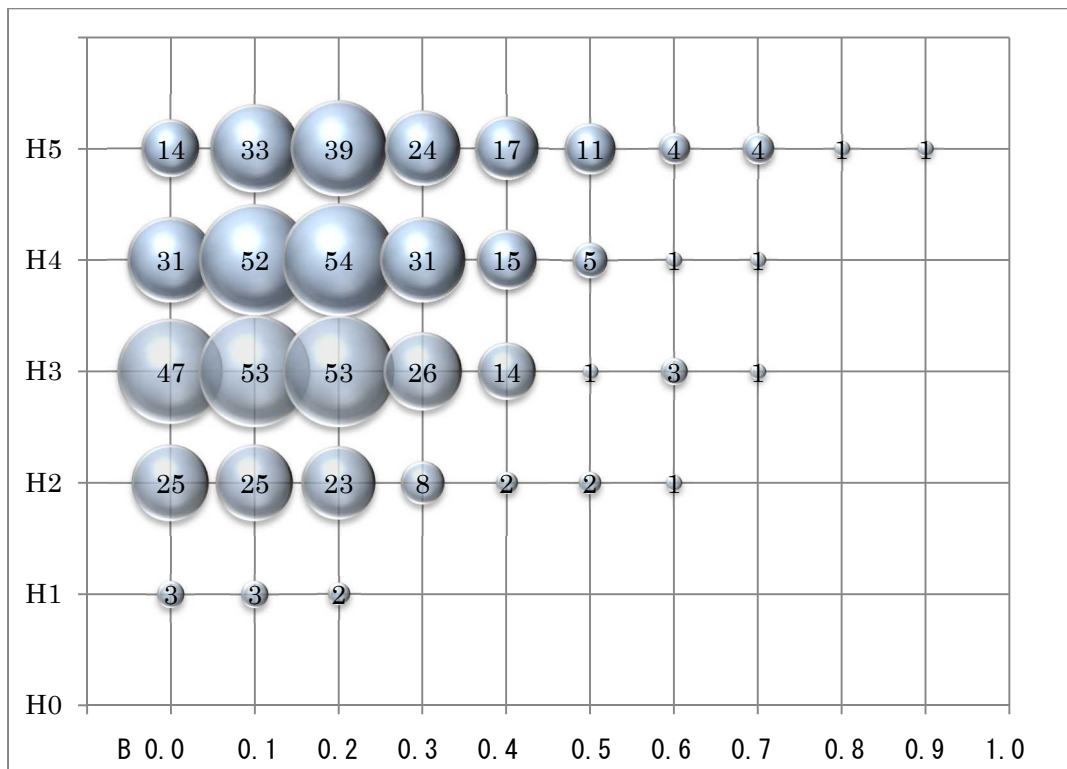


図 2-33 J-PlatPat の BLEU と人手評価

(b) J-PlatPat の NIST と人手評価 (「内容の伝達レベル」)

J-PlatPat の NIST と人手評価の Pearson の相関係数は 0.22 である。NIST 値は数値が大きいほど正解文の語の組み合わせが近いことを示すが NIST の値が 0.5~0.7 に人手評価の H1~H5 が分布しており、NIST の値と人手評価の値の対応がわからない。

	N0	N0.1	N0.2	N0.3	N0.4	N0.5	N0.6	N0.7	N0.8	N0.9	N1.0	N1.1	N1.2	N1.3	N1.4
H1	0	0	1	2	3	1	1	0	0	0	0	0	0	0	0
H2	0	4	5	10	14	16	18	11	4	2	1	1	0	0	0
H3	1	0	5	11	37	45	41	28	18	6	3	1	2	0	0
H4	0	0	3	12	23	36	42	36	23	8	5	0	1	1	0
H5	0	1	3	7	11	21	28	26	22	6	10	8	1	3	1

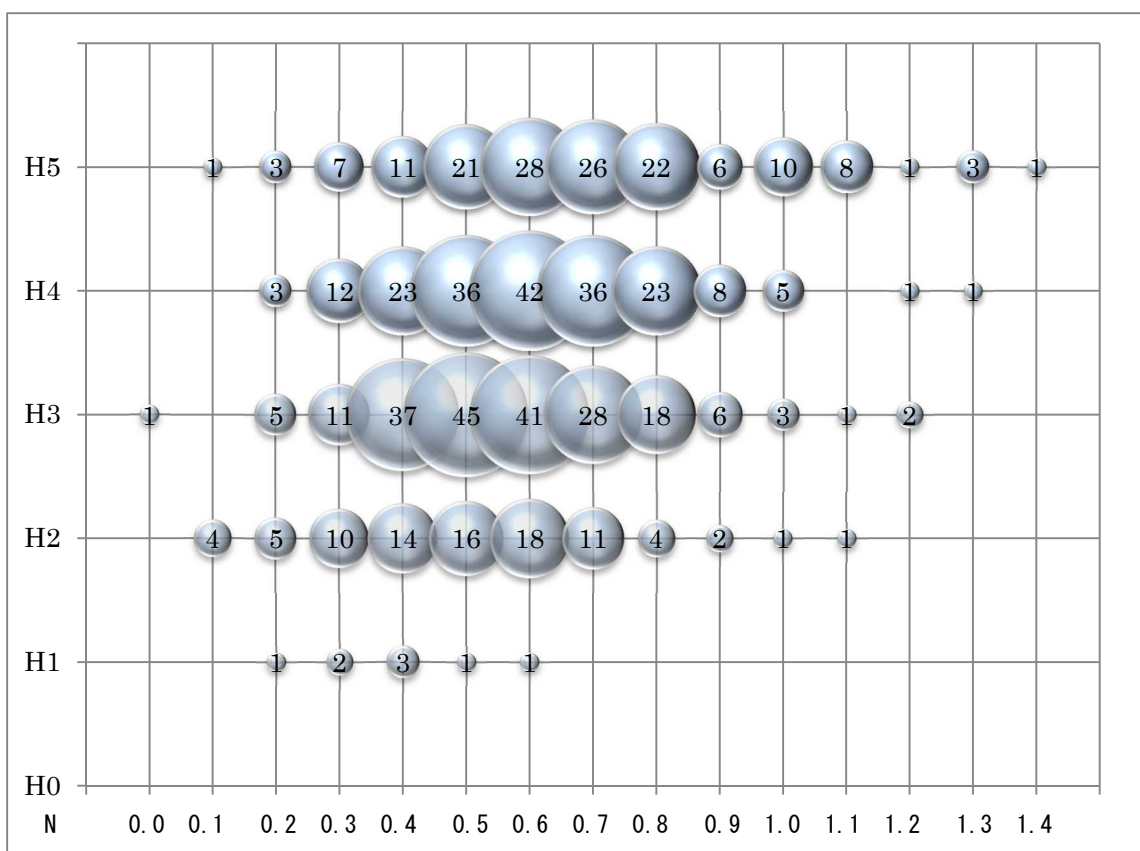


図 2-34 J-PlatPat の NIST と人手評価

(c) J-PlatPat の RIBES と人手評価（内容の伝達レベル）

J-PlatPat の RIBES と人手評価の Pearson の相関係数は 0.41 である。RIBES の値は 0.0～1.0 となり、1.0 に近いほど正解文の語の組み合わせと評価対象文が似ていることを示す。

RIBES の評価値が高くなるにつれて人手評価値も上がっているように見えるが、0.9 でも人手評価の H2、H3 を含むなど判定の精度は良くない。

	R0.0	R0.1	R0.2	R0.3	R0.4	R0.5	R0.6	R0.7	R0.8	R0.9	R1.0
H1	0	0	0	1	0	6	1	0	0	0	0
H2	0	0	0	3	10	18	27	23	4	1	0
H3	0	0	2	6	18	32	57	55	21	7	0
H4	0	0	0	6	10	17	33	50	63	11	0
H5	0	0	1	1	5	8	12	26	62	33	0

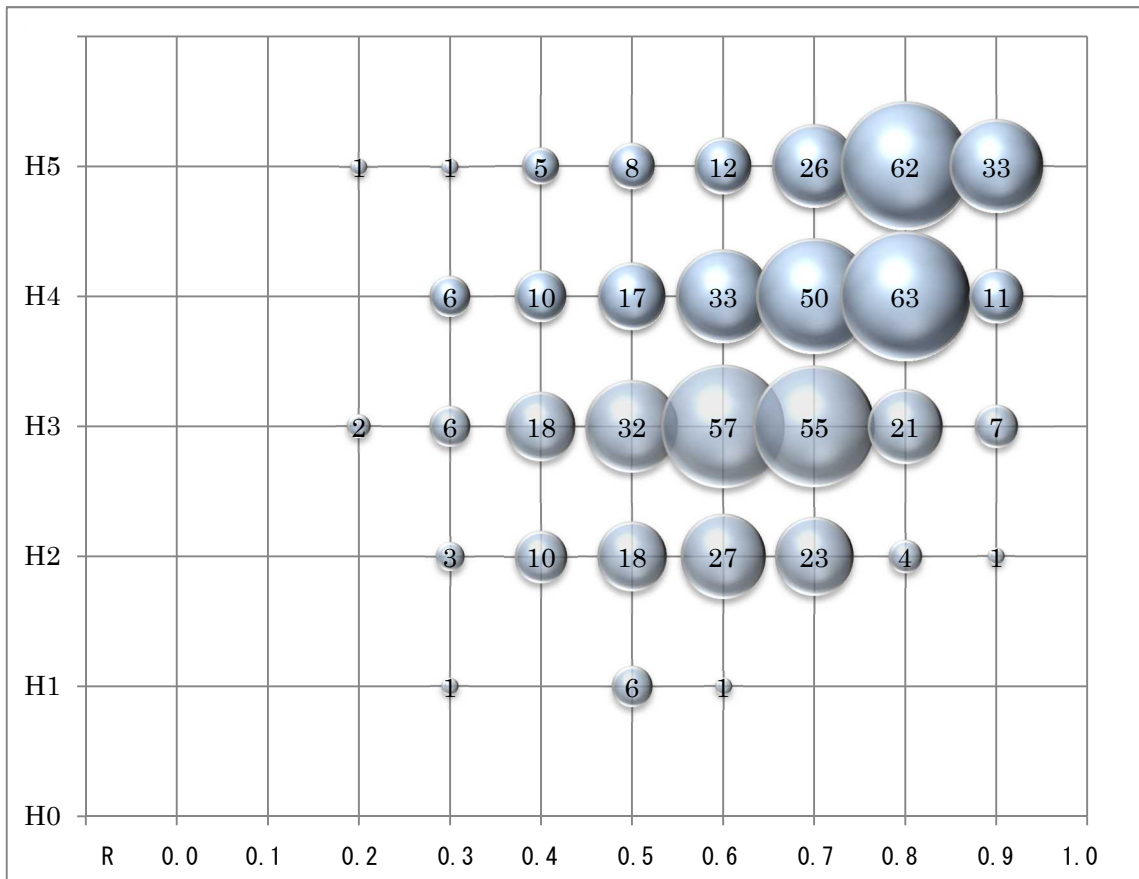


図 2-35 J-PlatPat の RIBES と人手評価

(2) 人手評価と自動評価の差の分析

前記(1)「相関係数の算出」では、人手評価と自動評価で強い相関があることが確認できなかった。ここでは、人手評価と自動評価との間で差の大きかったデータを例示しその内容を分析した。

まず、人手評価値が高く、自動評価値が低い例を下記表 2-18 に示す。

公開番号	JP2014240735A (分野 34)
内容の伝達レベル	4.5
自動評価	0.0605 (BLEU), 5.2394 (NIST), 0.286685 (RIBES)
原文	冷蔵庫の美観を保ちながら、ハンドルの取付けや取り外しを容易に行うことができる冷蔵庫を提供する。
基準翻訳文	The present invention is to provide a refrigerator such that a handle can be easily attached and detached while beauty of the refrigerator is maintained.
J-PlatPat	The refrigerator which can perform attachment and removal of a handle easily is provided maintaining the fine sight of a refrigerator.
備考	「内容の伝達レベル」は、美観には“fine sight”ではなく“fine appearance”が適切である以外内容の伝達に問題はないとし、評価値は4となった。 自動評価は、基準翻訳文と J-PlatPat の間で一致した単語が 12 単語（青字部分）となり、3 単語以上連続して一致した部分は存在しなかった。

表 2-18 人手評価値が高く、自動評価値が低い例

上記例では、J-PlatPat は原文の内容を伝えているものの、原文の“取り付けや取り外し”を基準翻訳文では受動的な訳になっているのに対して J-PlatPat では能動的な訳出のため、単語の一致具合が低くなったことにより、自動評価値が低くなったと考えられる。

上記例とは逆に人手評価値が低く、自動評価値が高かった例を表 2-19 に示す。

公開番号	JP2014240790A (分野 10)
内容の伝達レベル	2.5
自動評価	0.6298 (BLEU), 11.0477 (NIST), 0.875241 (RIBES)
原文	$P3'_{_1} [m] = P2_{_10} [m] + V2_{_10} [m/s] \times 10 [\mu s] \quad (9)$ そして、位置データ出力部 50 は、この式 (9) にて求められる計測時補正位置データのうち、data_req3 のデータ要求時点に最も時間的に近い計測時補正位置データを、上位システム 500 へと出力する。
基準翻訳文	$P3'_{_1} [m] = P2_{_10} [m] + V2_{_10} [m / s] \times 10 [\mu s] \quad (9)$ A position data output unit 50 outputs to the host system 500 a correction position data at measurement which is nearest in terms of time to the data request time of data_req3 from among the correction position data at measurement obtained by formula (9).
Google 翻訳	$P3'_{_1} [m] = P2_{_10} [m] + V2_{_10} [m / s] \times 10 [\mu s] \quad (9)$ Then, the position data output unit 50 of the measurement at the time the corrected position data obtained by the equation (9), the most temporally close measuring time correction position data to a data request time of data_req3, it will be output to the host system 500.
備考	「内容伝達のレベル」は、原文の構造が「位置データ出力部 50 は・・・位置データを、上位システム 500 へ出力する」であるが、Google 翻訳の訳文は主語「位置データ出力部 50」に係る述語が不明確となっており、文構造が崩れた訳出のため評価 1 となった。 自動評価は基準翻訳文と、Google 翻訳の間で多くの単語列が一致したため、高い評価値となった。

表 2-19 人手評価値が低く、自動評価値が高い例

この例では文の前半に数式があるため、基準翻訳文と Google 翻訳との間で多くの単語が一致し、3 単語以上の単語列の一致も 5 箇所存在するため自動評価値は高い値となった。一方人手評価である「内容の伝達レベル」は備考欄で説明したように訳文の構造が崩れているため内容が把握できず、低い評価となった。

3. 日英機械翻訳文の先行技術調査における有用性調査

日英機械翻訳文の先行技術調査における有用性調査は公開特許公報の全文日英機械翻訳文が、先行技術調査における英語のキーワード検索で発見可能かどうかを具体的に検証することで、全文日英機械翻訳文の先行技術調査における有用性を評価した。

3.1 調査の実施方法

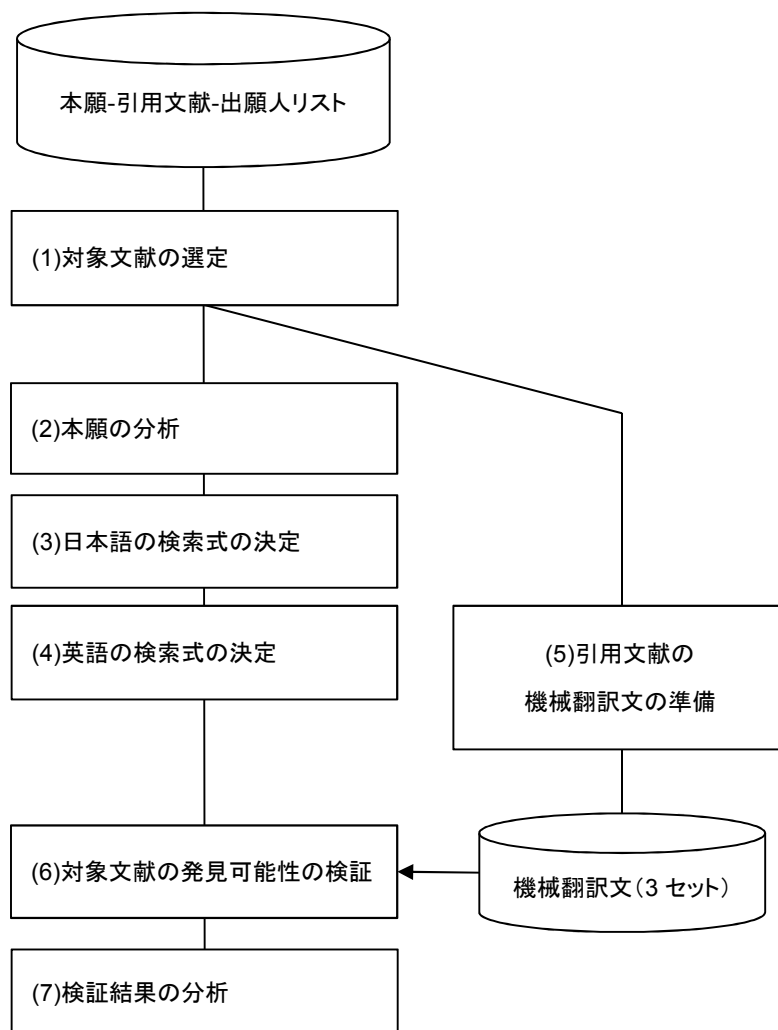


図 3-1 日英機械翻訳文の先行技術調査における有用性調査概要図

- (1) 対象文献の選定では、特許庁から貸与された「本願-引用文献-出願人リスト」を用いて調査対象とする引用文献を選定する。
- (2) 本願分析では、(1)で選定した引用文献を引用する拒絶理由通知書进行分析し、本願における請求項のどの発明特定事項に相当する先行技術として引用されたのかを

判断した。

- (3) 日本語の検索式の決定では、(2)で特定した本願の発明特定事項に対し、先行技術文献調査を行うことを想定して、日本語のキーワードを用いた検索式作成した。
- (4) 英語検索式の決定では、(3)で作成した検索式の日本語キーワードを英訳し、日本語の検索式に対する英語の検索式を決定した。
- (5) 引用文献の機械翻訳文の準備では、(1)で選定した引用文献の全てについて、英語版のJ-PlatPat及び特許庁が指定した2つの無料のオンライン機械翻訳サービスを用いて、全文の機械翻訳文を3セット用意した。
- (6) 対象文献の発見可能性の検証では、(4)で決定した英語の検索式によって、(5)で用意した全文の機械翻訳文が発見可能であるかを分析した。
- (7) 検証結果の分析では、(6)の検証結果を分析した。

3.1.1 調査条件・システム環境

(1) 技術分野

本調査の技術分野は、WIPO が作成した「Technology Concordance Table」に基づいた35の技術分野(項番5.1)を使用した。

(2) 機械翻訳文

評価対象とする機械翻訳文は、英語版 J-PlatPat サービス、espacenet サービス、及びみんなの自動翻訳を使用して取得した。

(3) 英語検索式検索ツール

前記図 3-1 の概要図(6)「対象文献の発見可能性の検証」では英語検索式を使った検索を行う。この検索システムとして当社 Japio-GPG/FX⁷ (以下、「GPG/FX」という。)の検索システムを利用した検索ツールを使用した。また、下記項番 3.2.5「英語検索式でのヒット件数調査」では、GooglePatents⁸を使用した。

3.1.2 対象文献の選定

日英機械翻訳文の先行技術調査における有用性調査での調査対象文献の選定は、日本国特許庁に対する特許出願 (以下、「本願」という。)の審査で引用された日本語の公開特許公報であって、PAJ が既に作成され、J-PlatPat で公開されている公開特許公報を対象に技術分野毎 2 文献を選定した。選定においては以下の (ア) から (エ) を選定条件とした。

⁷ <https://gpgfx.japio.or.jp/>

⁸ <https://patents.google.com/>

- (ア) 本願は、パテントファミリーとして英語の公報が発行されており、日本で審査の対象になった請求項（選定する公開公報を引用する拒絶理由と成る請求項）と同等の請求項が、英語の公報にも記載されていること。
- (イ) 本願に対する特許庁の審査において、選定する公開公報を引用する拒絶理由の対象となる請求項に対し、特許法 36 条に基づく拒絶理由が通知されていないこと。
- (ウ) 各技術分野で選定する 2 文献のうち、一方の文献は 1 回目の拒絶理由通知において特許法 29 条 1 項 3 号に基づく拒絶理由通知で引用されたものであり、他方の 1 文献は 1 回目の拒絶理由通知において特許法 29 条 2 項のみが通知され、その拒絶理由（他の文献との組み合わせにより進歩性を否定したものに限る）で引用されたものであること。
- (エ) 各技術分野で選定する 2 文献のうち、1 文献は国内出願人によるものであり、1 文献は外国出願人によるものであること。

表 3-1 調査対象文献選定の条件

3.1.3 本願の分析

項番 3.1.2 で選定した文献のそれぞれについて、当該文献を引用する拒絶理由通知書进行分析し、本願における請求項のどの発明特定事項に相当する先行技術として引用されたのかを判断し、選定した文献に対する発明特定事項を特定した。なお判断は、同一文献に対して 2 名が行った。

3.1.4 日本語の検索式の決定

項番 3.1.3 で特定した本願の発明特定事項（項番 3.1.2 で選定した 70 文献に対する 70 の発明特定事項）のそれぞれに対し、先行技術調査を行うことを想定して、項番 3.1.2 で選定した対象文献がヒットするような日本語キーワードを用いた検索式⁹を決定した。なお、作成した検索式は J-PlatPat の特許・実用新案テキスト検索にて検索を行い確認選定した文献がヒットすることを確認した。

⁹キーワードのみの検索式でヒット件数が多い場合、検索式に分類を加えヒット件数を調整した。

<<発明特定事項>>

【請求項 1】コーティング設備の副次面のための洗浄方法において、次の各ステップを含んでおり、すなわち、コーティングチャンバの副次面に付着防止層をコーティングに先行して塗布し、ドライアイス噴射またはＣＯ２スノー噴射洗浄法によって副次面をコーティングに後続して処理する方法。



<<日本語検索式>>

[コーティング/TX+コーティング/TX+被覆/TX]*[装置/TX+設備/TX]*[付着/TX+固着/TX]*[防止/TX+抑制/TX]*[ドライアイス/TX+ＣＯ２/TX]*[噴射/TX+ジェット/TX+ジェット/TX+吹きつけ/TX+吹き付け/TX+吹付け/TX+ブラスト/TX]*[洗浄/TX+除去/TX]*[真空/TX]*[面/TX+層/TX+基板/TX]*[C23C?/IP]

図 3-2 日本語検索式の決定例

3.1.5 英語の検索式の決定

項番 3.1.4 で決定した日本語のキーワード英訳し、項番 3.1.4 で決定した日本語の検索式に対応する英語の検索式を決定した。

■ 日本語検索式

[コーティング/TX+コーティング/TX+被覆/TX]*[装置/TX+設備/TX]*[付着/TX+固着/TX]*[防止/TX+抑制/TX]*[ドライアイス/TX+C O 2 /TX]*[噴射/TX+ジェット/TX+ジェット/TX+ジェット/TX+吹きつけ/TX+吹き付け/TX+吹付け/TX+ブラスト/TX]*[洗浄/TX+除去/TX]*[真空/TX]*[面/TX+層/TX+基板/TX]*[C23C?/IP]



■ 英語検索式（英語検索式の形式は調査対象の検索システム用）

coating (facility+facilities+apparatus +system+systems)
(adhesion+adhesive+depositing+deposition) (prevent*+suppression
+anti*+"anti-*"+inhibiting) ("dry ice"+CO2)
(jet+jetting+blast+blasting+spray+spraying) (clean+
cleaning+eliminate+remove) vacuum
(surface+surfaces+layer+substrate) (C23C?)

図 3-3 英語検索式決定例

なお日本語のキーワードの英訳には、本願の Patent ファミリーの公報や同義語の展開を行った。同義語の展開では下記図に示すような当社独自の用語展開ツールを利用した。

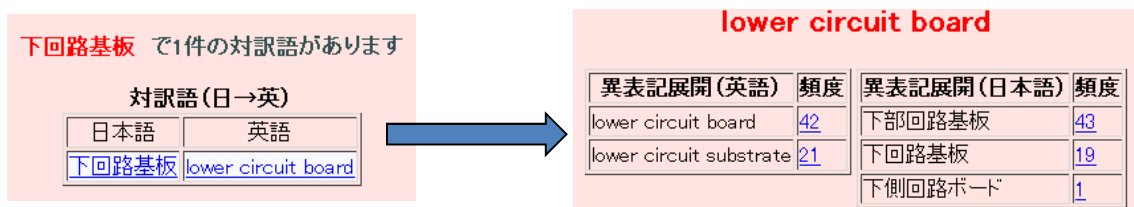


図 3-4 用語展開ツール

日本語検索式	英語検索式	備考
[コーテ イ ング/TX+コーテ イ ング/TX+被覆/TX]	coating	日本語表記ゆれ→1単語
[装置/TX+設備/TX]	(facility + facilit ies + apparatus + system + system s)	語尾展開
[付着/TX+固着/TX]	(adhesion + adhesive + deposit ing + deposit ion)	語尾展開
[防止/TX+抑制/TX]	(prevent* + suppression + anti* + "anti- * " + inhibiting)	語尾展開、接頭辞、接頭辞 それ以降のハイフネーション
[ドライアイス/TX+CO ₂ /TX]	("dry ice")	複合語をフレーズ検索指定
[噴射/TX+ジェット/TX+ジェット/TX+ジェット/TX+吹きつけ/TX+吹き付け/TX+吹付け/TX+ブラスト/TX]*	(jet + jet ting + blast + blast ing + spray + spray ing)	語尾展開
[洗浄/TX+除去/TX]*	(clean + clean ing + eliminate + remove)	同義語展開(除去→eliminate, remove)
[真空/TX]*	vacuum	
[面/TX+層/TX+基板/TX]*	(surface + surface s + layer + substrate)	語尾展開

表 3-2 英語検索式作成例

3.1.6 引用文献の機械翻訳文の準備

項番 3.1.2 で選定した 70 の文献の全てについて、英語版の J-PlatPat 及び特許庁が指定する 2 つの無料オンライン機械翻訳サービスを用いて全文の機械翻訳を 3 セット用意した (表 3-3)。

調査対象データ		J-PlatPat	Google 翻訳	みんなの翻訳
引用文献 (公開特許公報)	発明の名称	特許庁から貸与	Google 翻訳	みんなの自動翻訳@TexTraR
	要約	特許庁から貸与	Google 翻訳	
	請求項	J-PlatPat	espacenet	
	明細書	J-PlatPat	espacenet	

表 3-3 使用した機械翻訳システム (再掲)

3.1.7 対象文献の発見可能性の検証

対象文献の発見可能性の検証は、項番 3.1.5 で決定した英語の検索式によって、項番 3.1.6 で作成した 70 文献の全文の機械翻訳文が発見可能であるか、すなわち、検索式に含まれる英語のキーワードが、対応する文献の全文の機械翻訳文の中に存在しているかどうかを分析した。また、検証では比較検討のため日本語公報の要約文を英語に人翻訳した PAJ についての検証も合わせて行った。

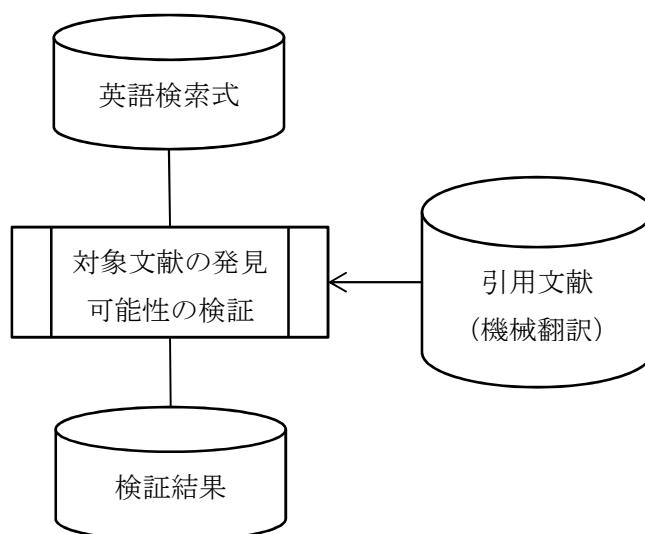


図 3-5 対象文献の発見可能性の検証概要図

3.1.8 検証結果の分析

項番 3.1.7「対象文献の発見可能性の検証」の検証結果で得た、3 セットの機械翻訳文及び PAJ の発見可能性と項番 2「日英機械翻訳の翻訳精度の調査」で得た人手による評価及び自動評価の結果との技術分野ごとの相関を分析した。また、拒絶理由の条文ごと、技術分野ごとの機械翻訳文及び、PAJ の今回選定した本願における先行技術文献としての有用性の差を分析した。

3.2 調査・分析結果

項番 3.1.7「対象文献の発見可能性の検証」にて検証した結果を分析した。分析は分野別、拒絶理由条文別に対象文献の発見可能性の傾向を分析するとともに、項番 2「日英機械翻訳の翻訳精度の調査」での調査・分析結果との関連性を分析した。また、PAJ についての先行技術調査を目的とした有用性を分析した。検証方法の課題等をまとめた。

3.2.1 分野別分析

分野別の集計は IPC8-Technology Concordance のフィールド単位(35 分野)だと分類数が細かすぎるため、IPC8-Technology Concordance のセクター単位で行った。

項番	セクター名	35 分類との対応
1	Electrical engineering (電気工学)	1～ 8
2	Instruments (機器)	9～13
3	Chemistry (化学)	14～24
4	Mechanical engineering (機械工学)	25～32
5	Other fields (その他)	33～35

表 3-4 IPC8-Technology Concordance(セクター単位) (再掲)

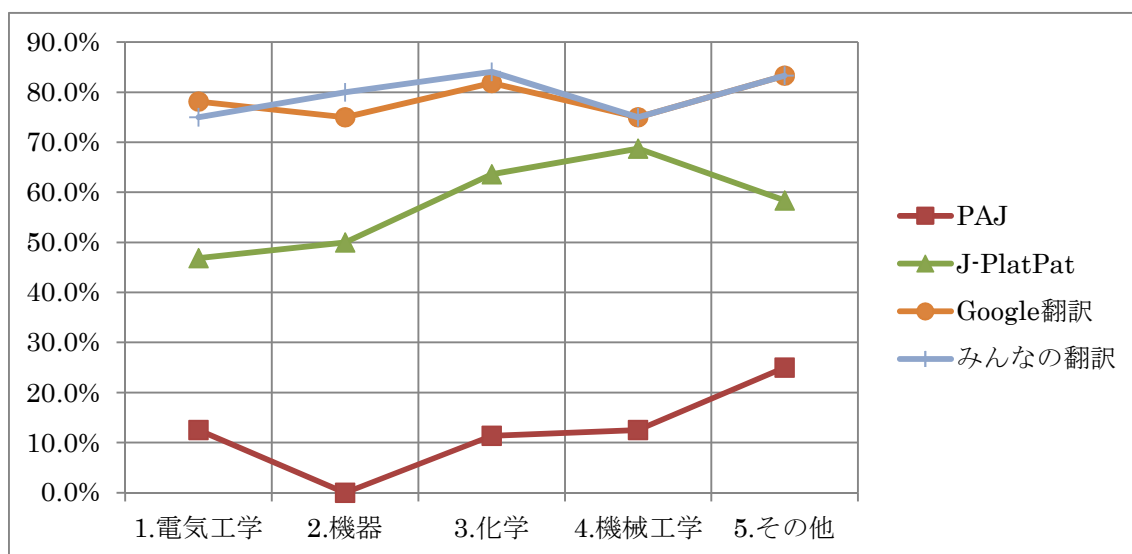


図 3-6 発見可能性検証結果（セクター別）

Google 翻訳とみんなの翻訳は引用文献を発見した割合が 80%前後で、分野間の差は少ない。J-PlatPat はセクター1(Electrical engineering)とセクター2(Instruments)の引用文献を発見した割合が 50%前後と低い。他のセクターも 60%から 70%となり、Google 翻訳とみんなの翻訳と比べ発見した割合が低い。また、PAJ はセクター5(Other fields)の発見率が最も高く 25%となり、セクター2(Instruments)が 0%、他のセクターが 10%強であった。

35 分野別に引用文献を発見できた件数の集計結果を下記表 3-5 に示す。

分野	PAJ	J-PlatPat	Google 翻訳	みんな の翻訳	分野	PAJ	J-PlatPat	Google 翻訳	みんな の翻訳
1	1	1	3	3	19	0	3	4	2
2	1	1	4	4	20	2	3	4	4
3	0	2	3	2	21	1	1	4	4
4	1	2	2	2	22	0	1	3	3
5	1	2	3	3	23	0	1	2	4
6	0	3	3	3	24	0	4	4	4
7	0	3	4	3	25	2	3	4	4
8	0	1	3	4	26	0	4	4	4
9	0	1	3	3	27	0	3	4	3
10	0	2	3	4	28	0	3	3	3
11	0	3	3	2	29	0	2	3	3
12	0	0	2	3	30	0	2	2	1
13	0	4	4	4	31	0	2	3	3
14	0	2	2	3	32	2	3	1	3
15	0	3	3	3	33	2	1	4	4
16	2	3	3	3	34	0	3	3	3
17	0	3	3	3	35	1	3	3	3
18	0	4	4	4	合計	16	82	110	111

表 3-5 分野別引用文献発見数

表 3-5 で機械翻訳の中で J-PlatPat だけが一件も見つけることができなかった分野 1(Control)の 4 式を詳細に調査した(表 3-6 から表 3-9)。その結果、いずれも品詞の違いによる技術用語の変化を原因とする微妙な違いで検索できなかったことが分かった。

このことは人手翻訳と機械翻訳どちらでも起こり得るので、英語のワード検索に置いて品詞の変化を吸収するような検索機能を備えることは、検索漏れの現象に有効と思われる。

引用文献公開番号	JP2008197850A
日本語検索式	[紙幣/TX+銀行券/TX+有価証券/TX]* [判別/TX+識別/TX]* [検査/TX]* [偽造/TX]* [カラー/TX+色/TX]* [波長/TX]
英語検索式	(bill + "bank note" + securities) (discrimination + identification) (inspection) (counterfeiting) (color) (wavelength)
原文（段落 0018）	第1種の方式は、該有価証券5の紙質を判別する方式である。
J-PlatPat	The system of the 1st type is a system which distinguishes the quality of paper of the negotiable securities 5.
備考	英語検索式"discrimination"（"判別"）に対して、J-PlatPat は distinguishes となり検索式と一致しなかった。
原文（段落 2）	目に付き易い偽造防止としては、材料、インク、図像部分があり、消費者が識別し易くなっており、目に見えない部分は肉眼では識別不能で、機器設備を利用し識別する必要がある。
J-PlatPat	There are material, ink, and an iconographic image portion, as anticounterfeit which stands out easily, it is easy to identify consumers and has become, and with the naked eye, the portion which is not visible is not discriminable, and needs to use and identify an equipment facility.
備考	英語検索式"counterfeiting"（"偽造"）に対して J-PlatPat では "anticounterfeit"となり検索式と一致しなかった。

表 3-6 検索式がヒットしなかった例（分野 12）

引用文献公開番号	JP2008197850A
日本語検索式	[銀行券/TX+金種/TX+有価証券/TX+紙幣/TX]* [識別/TX+区別/TX+認証/TX+検証/TX]* [発光ダイオード/TX+LED/TX]* [光電ダイオード/TX+フォトダイオード/TX+光センサ/TX]* [G07D7' /' ?/IP]
英語検索式	(banknote + securities + bill + "paper currency") (authentication + identification + discrimination + differentiation + verification) ("light emitting diode" + LED) ("optical diode" + "electro-optical diode" + photodiode + photo diode + "light sensor" + "optical sensor" + photosensor)
原文（請求項 1）	・ ・ ・、該移送通道の平面が貫通する光線を受ける第二レンズ と該第二レンズが受け取った光線を収集することができる光電 ダイオードを含み、 ・ ・ ・
J-PlatPat	・ ・ ・, and a <u>photoelectrical diode</u> which can collect beams of light which the second lens that receives a beam of light which a plane of the transfer adit penetrates, and the second lens received is included, ・ ・ ・
備考	英語検索式"optical diode" または "electro-optical diode" (" 光電ダイオード") に対して J-PlatPat では"photoelectrical diode"となり検索式と一致しなかった。

表 3-7 検索式がヒットしなかった例（分野 12）

引用文献公開番号	JP2003168178A
日本語検索式	[センサ/TX + モニタ/TX]* [各種プログラム/TX + 複数のプログラム/TX]* [選択/TX + 切替/TX + 切り替/TX + 切り換/TX + 切換/TX]* [分析/TX + 解析/TX]* [A61B5' /' ?/IP]
英語検索式	(sensor + monitor) ("a plurality of programs" + "multiple programs" + "multi programs") (select + selection + switch) (analysis)
原文（請求項 1）	【請求項 1】 各種プログラムを実行するためのCPUと、・・・
J-PlatPat	[Claim 1]A portable communication terminal comprising: CPU for running <u>various programs</u>
備考	英語検索式"a plurality of programs"（"各種プログラム"）に対して、J-PlatPat は"various programs"となり検索式と一致しなかった。
原文（段落 72）	このような場合には、アラームの形態を異ならせることで同一患者であっても異常の種類を判別することができる。また、発信した電子メールを解析することでも可能である。
J-PlatPat	In such a case, the kind of abnormalities can be distinguished even if it is the same patient in changing the form of alarm. It is also possible to <u>analyze</u> the sent E-mail.
備考	英語検索式 analysis("解析")に対して、J-PlatPat は analyze であったため、検索式と一致しなかった。

表 3-8 検索式がヒットしなかった例（分野 12）

引用文献公開番号	JP2003168178A
日本語検索式	[分析/TX+監視/TX+測定/TX+表示/TX]* [パラメータ/TX]* [プログラムモジュール/TX+拡張プログラム/TX]* [A61B5' /' ?/IP]
英語検索式	(analysis + monitoring + supervising + measurement + observation + display + indication) (parameter) ("program module" + "expansion program")
原文 (請求項 1)	【請求項 1】 各種プログラムを実行するためのCPUと、CPU の実行する基本的なプログラムを格納したプログラム記憶媒体 と、前記CPUが付加的に実行する拡張プログラムを格納した記 憶媒体を着脱自在に装着する記憶媒体装着手段と、・・・
J-PlatPat	[Claim 1]A portable communication terminal comprising: CPU for running various programs. A program storing medium which stored a fundamental program which CPU runs. Storage medium mounting which equips with a storage medium which stored an extended program which the aforementioned CPU runs additionally freely attachable/detachable.
備考	英語検索式"expansion program" ("拡張プログラム") に対して J-PlatPat が"extended program"となり、検索式と一致しなかつ た。

表 3-9 検索式がヒットしなかった例 (分野 12)

3.2.2 拒絶理由の条文別の分析

本調査で調査対象とした文献は、特許法 29 条 1 項 3 号に基づく拒絶理由通知で引用されたものと、特許法 29 条 2 項のみが通知された拒絶理由通知で引用された文献である（項番 3.1.2「対象文献の選定」を参照のこと）。拒絶理由条文別に発見可能性検証結果を集計した結果を以下に示す(図 3-7)。

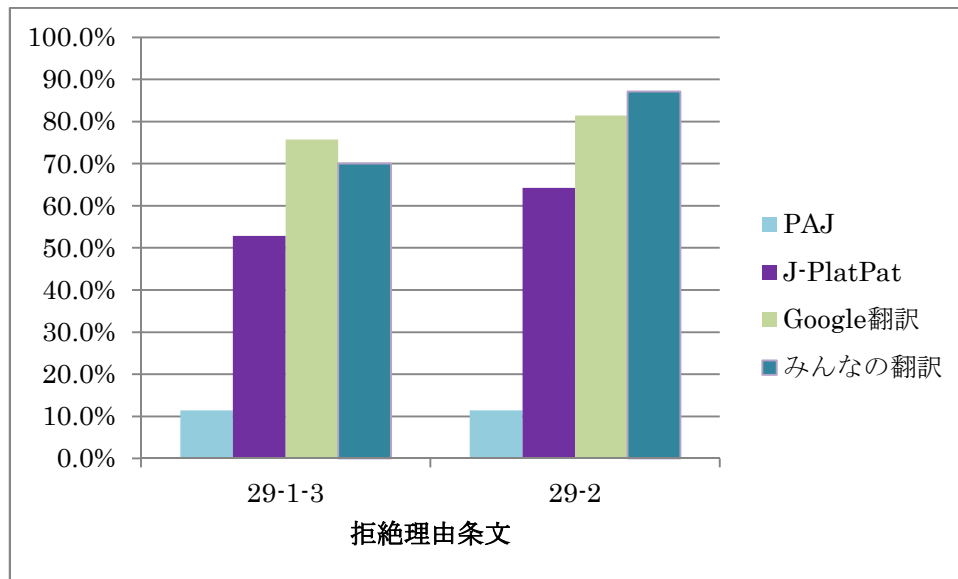


図 3-7 発見可能性検証結果（拒絶理由条文別）

拒絶理由条文別では 29 条 2 項の方が 29 条 1 項 3 号より若干高い傾向が見られた。また各機械翻訳及び PAJ 別に見ると、分類別結果と同様に Google 翻訳とみんなの翻訳の値が高い傾向がみられた。

3.2.3 日英機械翻訳の翻訳精度調査結果との相関

次に、項番 2「日英機械翻訳の翻訳精度の調査」の調査結果と「発見可能性検証結果」との関連を調査した。

まず「内容の伝達レベル」の評価結果と「発見可能性検証結果」を比較すると、「内容の伝達レベル」は項番 2.2.1「(1)「内容の伝達レベル」の評価結果」で示したように、翻訳精度は翻訳精度の高い順に J-PlatPat > みんなの翻訳 > Google 翻訳 の順となった。一方、「発見可能性検証結果」では発見可能性の高い順に みんなの翻訳 > Google 翻訳 > J-PlatPat となり「翻訳精度が高い＝発見可能性が高い」といった関連は見られなかった。

確認のため、J-PlatPat、Google 翻訳、みんなの翻訳の 35 技術分野について「内容の伝達レベル」と引用文献発見数の相関(表 3-10)を調べたところ、相関は見られなかった。

相関の調査対象	Pearson の 相関係数
J-PlatPat の内容の伝達レベルと引用文献発見数の相関	0.01
Google 翻訳の内容の伝達レベルと引用文献発見数の相関	0.01
みんなの翻訳の内容の伝達レベルと引用文献発見数の相関	0.12

表 3-10 35 技術分野の内容の伝達レベルと引用文献発見数の相関

次に「重要技術用語」の評価結果と比較すると、「重要技術用語」の評価結果は項番 2.2.1 「(2)「重要技術用語の翻訳精度」の評価結果」で示したように各機械翻訳間にほとんど差はなく、「発見可能性検証結果」との間に関連は見られなかった。

確認のため、J-PlatPat、Google 翻訳、みんなの翻訳の 35 技術分野について重要技術用語の翻訳精度と引用文献発見数の相関(表 3-11)を調べたところ、J-PlatPat、Google 翻訳についてはある程度の正の相関がみられた。一方で、みんなの翻訳の結果には相関がみられない。この結果から、平均値では薄まってしまうが、分野レベルで見ると重要技術用語の翻訳精度と引用文献発見数には弱い相関があると思われるが根拠は弱い。

相関の調査対象	Pearson の 相関係数
J-PlatPat の重要技術用語の翻訳精度と引用文献発見数の相関	0.21
Google 翻訳の重要技術用語の翻訳精度と引用文献発見数の相関	0.20
みんなの翻訳の重要技術用語の翻訳精度と引用文献発見数の相関	-0.09

表 3-11 35 技術分野の内容の重要技術用語の翻訳精度と
引用文献発見数の相関

最後に「自動評価」の評価結果と、比較すると項番 2.2.2 「自動評価結果の分析」で示したように、自動評価手法「BLEU」及び「NIST」の評価結果が翻訳精度の高いほうから Google 翻訳、みんなの翻訳、J-PlatPat の順となり、「発見可能性検証結果」と似た傾向が見られた。

確認のため、自動評価の BLEU と RIBES の評価値と引用文献発見数の相関を調べたところ、相関は見られなかった。この結果から自動評価の BLEU と RIBES の評価値とは相関が無いと思われる。これは、BLEU の評価値が人手評価の値と相関が見られなかったことから妥当な結果と思われる。

表 3-12 35 技術分野の内容の自動評価値と引用文献発見数の相関

3.2.4 PAJ の有用性分析

項番 3.1.7「対象文献の発見可能性の検証」では、3 つの機械翻訳の他、PAJ についても発見可能性の検証を行った。検証の結果は前記表 3-5 で示したように調査対象とした 140 式中 16 式がヒットした。これは、各機械翻訳が 80 以上の式でヒットしたのに比べかなり少ない結果となった。

この結果の要因として、まず各機械翻訳は公開公報全文が検索対象であるに対し、PAJ は要約を人手で英語に翻訳したというデータの量の差が考えられる。例えば、検索式で指定した用語が要約中には存在せず、請求項や詳細な説明のみに存在するような場合 PAJ はヒットしなくなる。ちなみに本調査で使用したデータ文字数を調べると、公報全文の機械翻訳はおよそ 47,000 文字/文献であるのに対し、PAJ はおよそ 1,000 文字/文献であった。

また、少数ながら PAJ でヒットし、機械翻訳でヒットしないものも存在した。140 式の内、J-PlatPat でヒットせず PAJ でヒットしたのが 9 式で最も多く、Google 翻訳でヒットせず PAJ でヒットしたのが 3 式、みんなの翻訳でヒットせず PAJ でヒットしたのが 1 式であった。

この中の 1 件の PAJ と J-PlatPat の翻訳を確認したところ下記表 3-13 の例に示すように、語尾変化の差により機械翻訳がヒットしなかったためであり、PAJ の翻訳品質によって PAJ のみがヒットしたのではなかった。

引用文献公開番号 JP2007039118A	
日本語検索式	[錠剤/TX]* [搬送ドラム/TX+回転ドラム/TX]* [カプセル/TX+溝/TX+孔/TX+穴/TX]* [充填/TX+収納/TX+収容/TX]
日本語原文	【解決手段】 錠剤供給部 3 2 から搬送ドラム 3 1 の外周面に形成されたポケット 3 1 1 内に扁平錠剤 t を起立状態で収容して搬送ドラムの下側に搬送し、ポケットから搬送ドラムの下方のガイドシュート 3 3 1 へ起立状態で落下させ、シュートの曲線軌道に沿って平伏状態で姿勢を変換して平伏状態で計数ディスク 2 に形成された計数孔 2 1 に投入し、シャッター板 4 を開いて、計数孔の下方に配置されたカプセルボディーに扁平錠剤を平伏状態で充填するカプセルへの錠剤充填装置。 【効果】 扁平錠剤をカプセルに充填する際の充填不良を可及的に防止して、所定個数の扁平錠剤を計数して平伏状態で確実に充填することができ、カプセルへの錠剤充填操作における充填不良の発生が極めて少なく信頼性が高い。 【選択図】 図 2
英語検索式	(tablet) ("conveyance drum" + "conveying drum" + "rotary drum" + "rotating drum") (capsule + hole + opening) (filling + filler + storage + housing + storing + accommodation)
PAJ	PROBLEM TO BE SOLVED: To provide a reliable apparatus for filling a capsule with tablets capable of minimizing failures in filling the capsule with flat tablets and counting a predetermined number of the flat tablets for surely putting them in a lying posture, with the failures extremely reduced in filling the capsule with the tablets. SOLUTION: The apparatus for filling the capsule with the tablets is constituted to accommodate the flat tablets t in a standing posture in a pocket 311 formed on an outer peripheral surface of a conveyance drum 31 from a tablet supply part 32 to carry them to below the drum, drop them from the pocket into a guide chute 331 below the drum in the standing posture, change the posture of the tablets into a lying posture along a curved track of the chute for putting them in the lying posture into a counting hole 21 formed on a counting disk 2, open a shutter plate 4, and fill a capsule body provided below the counting hole with the flat tablets in the lying posture. <P>COPYRIGHT: (C)2007, JPO&INPIT

J-PlatPat	[Document Name]Abstract [Abstract] [Means for Solution]<u>Store</u> the flat tablet t by a standing state in the pocket 311 formed in an outer periphe+B20ral surface of the <u>conveyance drums</u> 31 from the <u>tablet</u> feed zone 32, and it conveys below conveyance drums, It is made to fall from a pocket by a standing state to the lower guide chute 331 of conveyance drums. converting an attitude to a prostration state along a curved orbit of a shot -- it is calculation in a prostration state -- calculation formed in the disk 2 -- supplying to the <u>hole</u> 21 and opening the shutter plate 4 -- calculation -- a tablet <u>filling</u> apparatus to a <u>capsule</u> which fills up a flat tablet with a prostration state into a capsule body arranged under the hole. [Effect]The short shot at the time of filling up a capsule with a flat tablet is prevented as much as possible, the flat tablet of a prescribed number can be calculated and it can fill up with a prostration state reliably, and there is very little generating of the short shot in the tablet filling operation to a capsule, and it is reliable. [Chosen drawing] Fig.2
備考	・英語検索式のキーワード“conveyance drum”（搬送ドラム）は、J-PlatPat では“conveyance drum<u>s</u>”となり、複数形の s が一致しなかったため、検索式全体でヒットしなかった。 ・英語検索式のキーワード“storing”（収容）は J-PlatPat では“stor”となり、語尾が一致しなかった。 ・英語検索式のキーワード“accommodation”（収容）は、PAJ では“accommodate”となり、語尾が一致しなかった。

表 3-13 PAJ がヒットし、J-PlatPat がヒットしなかった例

3.2.5 英語検索式でのヒット件数調査

項番 3.1.4 から項番 3.1.6 で日本語検索式によるヒット件数が妥当な件数（200～500）となるように日本語検索式を作成し、その日本語検索式を英語検索式に翻訳し、翻訳された英語検索式で各種機械翻訳結果の英文が検索可能であるか調査した。この方法で英語検索式で検索できることが分かるが、英語のヒット件数が大幅に増加してスクリーニングに困る事態が考えられる。そこで、翻訳された英語検索式で機械翻訳された英語データを実際に検索したときのヒット件数と、日本語検索式のヒット件数の違いを調査した。本調査は発展的な調査である。

なお、GooglePatents についてはデータ蓄積の範囲、表 3-17 で示した「Similar Documents」のタイトルでもヒットする等、正確な仕様が不明であるため、調査した事実を示すにとどめる。本章は明確な結論を出せない。

(1) 調査条件

機械翻訳を英語検索式で検索した場合、ルールベースの機械翻訳結果と統計ベースの機械翻訳結果でヒット件数の違いが考えられる。理由はルールベース翻訳が辞書に基づいて統制された英語を生成するのに対して、統計ベース翻訳は事例に基づいた周辺の語に影響された非統制の英語を生成する可能性があるためである。

そこで、本調査では検索式数件についてルールベースの機械翻訳結果と統計ベースの機械翻訳結果の違いを調査した。具体的には選択した同じ英語検索式を使って 2 つの英語検索システム（GooglePatents および GPG/FX）でヒット件数を取得した。そして、英語検索式でのヒット件数と日本語検索式でのヒット件数を比較・分析した。

GooglePatents は統計ベースの機械翻訳データを使用した検索システムの代表、GPG/FX はルールベースの機械翻訳データを使用した検索システムの代表として調査対象とした。なお、GooglePatents は IPC 検索ができないため、IPC を含む検索式は調査対象から除外した。なお、GooglePatents は CPC の検索が可能である。

調査は項番 3.1.5 「英語の検索式の決定」で作成した検索式の中の IPC 分類を指定していない式に限定し、さらに技術分野に偏りがないように 34 式を選択した。また、調査で使用する検索システム間のデータ蓄積範囲を揃えるため、各検索式に公開日範囲を追加した。

(2) 調査結果（日本語検索式のヒット件数との比較）

前記(1)「調査条件」で示した検索式を使い英語検索式で検索をした結果を図 3-8 に示す。

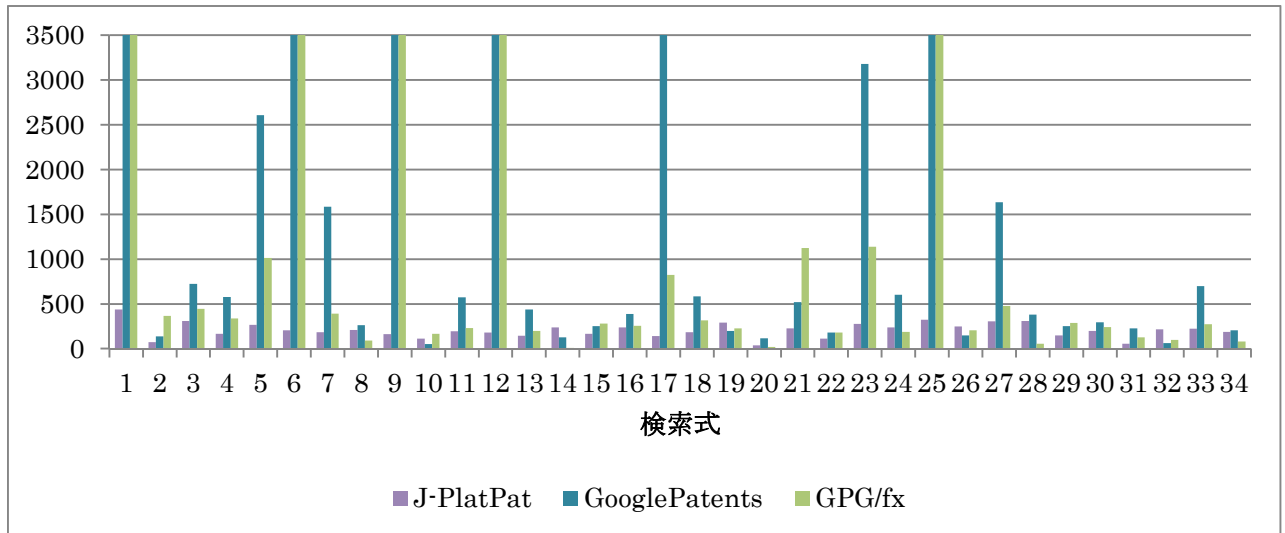


図 3-8 日本語検索式と英語検索式のヒット件数
(縦軸のヒット件数 3,500 件以上はグラフでは省略)

調査結果から、日本語検索式のヒット件数に比べ、英語検索式によるヒット件数が多い傾向が見られた。特に、上図の検索 1、6、9、12、17、25 番目の検索式では日本語検索式のヒット件数が数百件であるのに対して、英語検索式のヒット件数が 3,500 件を超える結果となった。調査した 34 式の内、GooglePatents のヒット件数よりも GPG/fx のヒット件数が多かった式は 8 式であった。

上記図 3-8 で特に件数の差が大きかった検索式 1 番目と 12 番目の検索式及びヒット件数についてヒット件数の差の要因を調査した。

日本語検索式 1	[プリンター /TX+印刷機/TX]*[スクイジーローラ/TX+除去ローラ/TX+ 清掃ローラ /TX+ 清浄ローラー /TX+洗浄ローラ/TX]
英語検索式 1	("printer"+"printing apparatus"+"printing device"+"printing machine") ("squeegee roller"+"removing roller"+"removal roller"+" cleaning roller "+"washing roller")
J-PlatPat (日本語)	438
GooglePatents	15,526
GPG/FX	13,166

表 3-14 検索式 1 のヒット件数

英語検索式 1 でヒットし、日本語検索式 1 でヒットしない文献の内容を確認すると、用語”printer”（プリンター）や “cleaning roller”（クリーニングローター）でヒットしたものが多く見られた。これらの文献の原文を確認すると、原文の表記が”プリンタ”（長音なし）や “クリーニングローラ”（長音なし）となっているものが多く、これらの文献が英語検索式で多くヒットしたと考えられる。なお、日本語検索式に”プリンタ”、 ”クリーニングローラ”を追加し再度 J-PlatPat で検索すると、11,003 件となり、このことからこれらの用語が件数差の主な要因と考えられる。

日本語検索式 12	[染毛剤/TX+ 髪 /TX+染毛料/TX+毛染剤/TX]*[容器/TX+装置/TX]*[開封 /TX]*[出口/TX+注ぎ口/TX]
英語検索式 12	(hair +”hair dye”+”dyeing agent”) (container+device+vessel+chamber+unit+apparatus+machine+mechanism) (opening) (outlet +exit+spout+tap+”opening spout”)
J-PlatPat(日本語)	181
GooglePatents	19, 892
GPG/FX	15, 267

表 3-15 検索式 12 のヒット件数

英語検索式 12 でヒットし、日本語検索式 12 でヒットしない文献の内容を確認したところ、下記(表 3-16)のように用語”hair”（髪）や、”opening”（開封）でヒットする部分の日本語原文を確認すると”毛”や”開口”となっており、これら用語が件数差の主な要因と考えられる。なお、日本語検索式に”毛”、 ”開口”を加え再度検索すると、24, 294 件ヒットした。ことからこれらの用語が件数差の主な要因と考えられる。

つまり、日本語検索式と英語検索式のヒット件数の違いは、日本語検索式から英語検索式に変換する過程で英語に翻訳した際、一般的な用語に翻訳してしまったケースが主な要因と考えられる。

This cleaning robot body housing by <u>opening</u> the suction port on the bottom surface to the interior of the electric blower and dust collection unit performs the cleaning on the floor and self-propelled この掃除ロボットは吸込口を底面に<u>開口</u>して電動送風機及び集塵部を内装する本体筐体が床面上を自走して掃除を行う。 <u>opening</u> the outlet port 35a at the tip are derived. 流出口 3 5 a を<u>開口</u>して
the bypass passage 20 for introducing the combustion exhaust gas G is passed through the cylindrical portion 22 <u>opening</u> is surrounded by the throttle section 23, バイパス流路 2 0 が筒部 2 2 を挿通すると共に先端<u>開口</u>部が絞り部 2 3 に圍繞されており with washing <u>hair</u> in deposits F. 洗浄<u>毛</u>を付着物 F に

表 3-16 英語検索式 12 でヒットし、日本語検索式 12 でヒットしない文献

(3) 調査結果（機械翻訳の違いによるヒット件数の差）

英語検索式の検索は前記(1)の調査条件にあるように GooglePatents と GPG/FX を使用した。GooglePatents は統計ベースの機械翻訳のデータを使用し、GPG/FX は特許庁の機械翻訳システムと同じルールベースの機械翻訳のデータを使用している。機械翻訳の種類の違いがヒット件数に影響するかの調査をした。対象はヒット件数が少なく、GooglePatents と GPG/FX のヒット件数差が少ないものを対象とした。これは前記(1)で分析したようにヒット件数が非常に多いものや、2 つのヒット件数の差が大きいものはノイズが多く含まれていると考え対象外とし GooglePatents と GPG/FX のヒット件数の差が最も少ない式について調査を行った。調査対象の検索式及びヒット件数を表 3-17 に示す。

日本語検索式	[エネルギー/TX]*[再生/TX+回収/TX+回生/TX]*[可変容量/TX]*[油圧/TX]*[ブームシリンダ/TX]*[アームシリンダ/TX]
英語検索式(*1)	(energy) (reproduction+recovery+regenerative) ("variable capacity") ("oil pressure"+hydraulic) ("boom cylinder") ("arm cylinder")
GooglePatents ヒット件数(*2)	61
GPG/FX ヒット件数(*3)	98

表 3-17 比較調査対象検索式及びヒット件数

(*1)GooglePatents と GPG/FX の蓄積期間を合わせるため、上記英語検索式に公開日 2014 年 12 月 31 日以前を指定した。

(*2)(*3)上記表のヒット件数は、ヒット結果の比較対象を公開公報のみに合わせるためヒット結果から国際公開公報(2 件)、公表公報(2 件)を除外した件数となる。

次に GooglePatents と GPG/FX それぞれヒットした公報の番号を突き合わせ比較を行った(図 3-9)。

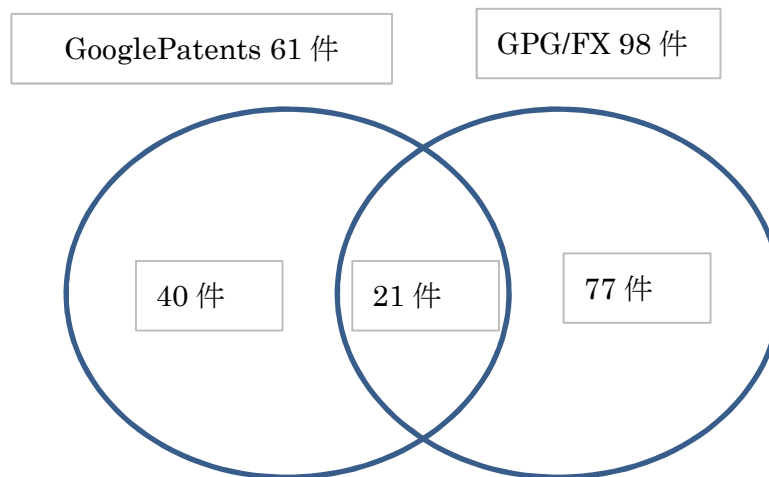


図 3-9 ヒットした番号の比較

ヒットした番号の比較結果のうち、GooglePatents のみでヒットした 40 件の中から 2 件、GPG/FX のみでヒットした 77 件の中から 2 件を抽出し、他方の検索式でヒットしなかった要因を調査した。調査結果を下記表 3-18 から表 3-21 に示す。

今回の調査においては機械翻訳の方式の違いを要因とする違いは見られなかった。

公開番号	JP2014145387A
原文	【0002】油圧ショベルなどの油圧駆動の作業機械に適用されたエネルギー再生回路の一例として、コントロール弁とタンクとの間に設けた戻り流体通路中に、可変容量型油圧モータなどの流体圧モータをインラインで設置・・・
GooglePatents の機械翻訳文	As an example of the energy recovery circuit is applied to a working machine of a hydraulic drive, such as a hydraulic excavator, it is placed in the return fluid passage in which is disposed between the control valve and the tank, a fluid pressure motor, such ・・・
GPG/FX の機械翻訳文	[0002] [as an example of the energy reflex circuit applied to hydraulic driven work machines, such as a hydraulic excavator,]Into the return fluid passage provided between the control valve and the tank, place fluid pressure motors, such as a variable displacement hydraulic motor, with in-line one, and there is through a decelerator. ・・・
備考	GPG/FX の機械翻訳文は“再生”に対する訳語“reflex”が英語検索式に該当しないため、ヒットしなかった。

表 3-18 GooglePatents のみでヒットした文献 1

公開番号	JP2014126126A																											
GooglePatents のヒット箇所	Similar Documents <table><thead><tr><th>Publication</th><th>Publication Date</th><th>Title</th></tr></thead><tbody><tr><td>JP2010190261A</td><td>2010-09-02</td><td>Hydraulic control device for working machine, and working machine equip</td></tr><tr><td>CN1347818A</td><td>2002-05-08</td><td>Speed control apparatus for working vehicle and its control method</td></tr><tr><td>US7356991B2</td><td>2008-04-15</td><td>Hydraulic control device of an excavator with improved loading performanc</td></tr><tr><td>US7178333B2</td><td>2007-02-20</td><td>Hydraulic control system for hydraulic excavator</td></tr><tr><td>US6282890B1</td><td>2001-09-04</td><td>Hydraulic circuit for construction machines</td></tr><tr><td>JP2010084888A</td><td>2010-04-15</td><td>Power regenerating mechanism of hydraulic working machine</td></tr><tr><td>CN1697933A</td><td>2005-11-16</td><td>Oil pressure circuit for working machines</td></tr><tr><td>US20020108486A1</td><td>2002-08-15</td><td>Hydraulic recovery system for construction machine and construction mac</td></tr></tbody></table>	Publication	Publication Date	Title	JP2010190261A	2010-09-02	Hydraulic control device for working machine, and working machine equip	CN1347818A	2002-05-08	Speed control apparatus for working vehicle and its control method	US7356991B2	2008-04-15	Hydraulic control device of an excavator with improved loading performanc	US7178333B2	2007-02-20	Hydraulic control system for hydraulic excavator	US6282890B1	2001-09-04	Hydraulic circuit for construction machines	JP2010084888A	2010-04-15	Power regenerating mechanism of hydraulic working machine	CN1697933A	2005-11-16	Oil pressure circuit for working machines	US20020108486A1	2002-08-15	Hydraulic recovery system for construction machine and construction mac
Publication	Publication Date	Title																										
JP2010190261A	2010-09-02	Hydraulic control device for working machine, and working machine equip																										
CN1347818A	2002-05-08	Speed control apparatus for working vehicle and its control method																										
US7356991B2	2008-04-15	Hydraulic control device of an excavator with improved loading performanc																										
US7178333B2	2007-02-20	Hydraulic control system for hydraulic excavator																										
US6282890B1	2001-09-04	Hydraulic circuit for construction machines																										
JP2010084888A	2010-04-15	Power regenerating mechanism of hydraulic working machine																										
CN1697933A	2005-11-16	Oil pressure circuit for working machines																										
US20020108486A1	2002-08-15	Hydraulic recovery system for construction machine and construction mac																										
備考	GooglePatents は公報以外に「Similar Documents」という項目があり、この中の記載に検索用語”recovery”が含まれていたため、ヒットした。																											

表 3-19 GooglePatents のみでヒットした文献 2

公開番号	JP2014145387A
原文	【0002】特許文献1には、作動油を吐出してアクチュエータを駆動する可変容量型のメインポンプと、メインポンプの出力をアシストするように作動油を吐出するサブポンプと、・・・
GooglePatents の機械翻訳文	Patent Document 1, a <u>variable displacement</u> main pump for driving an actuator by discharging the hydraulic oil,・・・
GPG/FX の機械翻訳文	[0002] The main pump of the <u>variable-capacity</u> form which discharges a hydraulic oil in PTL1 and drives an actuator to it, The hybrid construction machine equipped with the sub-pump which discharges a hydraulic oil・・・
備考	GooglePatents の機械翻訳文は“可変容量”に対する訳語“variable displacement”が英語検索式に該当しないため、ヒットしなかった。

表 3-20 GPG/FX のみでヒットした文献 1

公開番号	JP2014009525A
原文	【0051】(2)回収電力の蓄電制御 車体コントローラ11は、旋回制動時にインバータ13を制御して旋回電動機16を発電機として動作させる発電制御を行い、旋回電動機16から電気エネルギーを回収するとともに、回収した電気エネルギーをバッテリー15に蓄える制御を行う。
GooglePatents の機械翻訳文	(2) electrical storage control body controller 11 for <u>recovering</u> electric power, performs power generation control to the turning electric motor 16 by controlling the inverter 13 during the turning braking is operated as a generator, as well as collecting electrical energy from the rotation motor 16, and the recovered electric It does control for storing energy in the battery 15.
GPG/FX の機械翻訳文	[0051] (2) Accumulation-of-electricity control of <u>recovery</u> electric power Body controller 11 performs control which stores the recovered electric energy in battery 15 while it performs power generation control which controls inverter 13 at the time of circling braking, and operates circling electric motor 16 as a dynamo and recovers electric energies from circling electric motor 16.
備考	GooglePatents の機械翻訳文は“回収”に対する訳語“recovering”が英語検索式に該当しないため、ヒットしなかった。

表 3-21 GPG/FX のみでヒットした文献 2

また、GooglePatents のみでヒットした機械翻訳の訳語を見ると同一文献内の同一用語に複数の訳語が出力されているものが存在した(表 3-22)。

公開番号	JP2006125566A
原文	【0024】 請求項9記載の発明は、上記両方向ポンプモータは <u>可変容量</u> 式である建設機械の油圧シリンダ電動装置を提供する。
GooglePatents の機械翻訳文	Invention of claim 9 wherein, said bi-directional pump motor to provide a hydraulic cylinder electric system for a construction machine is a <u>variable displacement</u> type.
原文	【0036】 請求項9記載の発明は、両方向ポンプモータは <u>可変容量</u> 式であるので、容量を上昇させて回転数を高くすることにより、油圧シリンダ伸長操作時は油圧シリンダの動作速度を速くすることができる。
GooglePatents の機械翻訳文	Invention of claim 9 wherein the bi-directional pump motor is therefore a <u>variable capacity</u> type, by increasing the rotational speed to thereby increase the capacity, when the hydraulic cylinder stretching operation can be faster the operating speed of the hydraulic cylinder.
備考	“可変容量”の訳語が段落24では“ <u>variable displacement</u> ”と訳され、段落36では“ <u>variable capacity</u> ”と訳されている。

表 3-22 複数訳語例(GooglePatents)

対訳コーパスに基づく統計ベースの翻訳システムでは対訳コーパス内で一貫した用語統一が行えないことがあり、このような訳語の振れが生じやすい。この訳語の振れは検索の観点では検索の網羅性（リコール）を上げる方向働くが、一方で検索結果にノイズを含んでしまい検索精度（プレジジョン）を下げる方向に働くことがある。

4. 問題点・課題等

本章では特許庁評価手順を日英翻訳の評価に適用するにあたっての修正点、評価の作業負担・課題(判断が困難であった点等)、及び人手による評価と自動評価の組み合わせの手法等について提案する。

4.1 内容の伝達レベルの評価作業に関する問題点・課題

項番 2 の「日英機械翻訳の翻訳精度の調査」の「内容の伝達レベル」の評価は複数の評価者により評価を行った。評価作業は評価内容のズレが生じないように事前に意識合わせを行ったが、評価結果を確認すると評価者間のズレが生じていたため再度評価した。このような課題に対して、下記付録 5.3「評価者間の評価の一致度合いを調べる指標」に示すような方法を活用して、ズレが大きな評価者をあぶりだすことが望ましいが、評価者が二人の場合、どちらの評価が正しいのか判定する必要がある、コスト増と所要時間の増大を招く。作業期間と予算の裏付けが必要である。

4.2 重要技術用語の翻訳精度の評価作業に関する問題点・課題

重要技術用語の翻訳精度において、頻度は多くないが二つの課題があった。一つは日本語の技術用語が複合語名詞であるのに対して、英語では of を含むフレーズとして翻訳されていること、二つ目は同一の日本語の技術用語が一文の中で複数回現れた場合、その英語訳が一つに定まらない場合である。以下に例で説明する。

日本語の技術用語が複合語名詞、英語は of を含むフレーズの場合、抽出した重要技術用語表現が基準翻訳文中に存在しない場合がある。そのような場合は、基準翻訳文の英語訳はあまり参考に出来ないので、日本語の重要技術用語「排出量」に対して、機械翻訳文中の訳語が正しいかどうかを判定する必要がある。

公開番号	JP2014240279A
原文	C P U 1 1 が、燃料センサにより検出された燃料消費量から車両から排出される C O 2 の単位時間当たりの排出量である瞬時排出量を算出し、算出した瞬時排出量と閾値とを比較して、その比較結果に基づいて瞬時排出量を評価する。
重要技術用語	排出量
基準翻訳文	A CPU 11 calculates an amount of instantaneous emission which is a discharge amount per unit time of CO2 discharged from a vehicle on the basis of a fuel consumption amount detected by a fuel sensor, compares the calculated amount of instantaneous emission with a threshold, and evaluates the amount of instantaneous emission on the basis of comparison result.
重要技術用語	amount of emission

J-PlatPat	CPU11 computes an instant exhaust amount which is an exhaust amount per unit time of CO2 discharged from vehicles from fuel consumption detected by a fuel sensor, compares an instant exhaust amount and a threshold value which were computed, and evaluates an instant exhaust amount based on the comparison result.
重要技術用語	exhaust amount
技術用語評価	A
Google 翻訳	CPU11 is, calculates the instantaneous emissions are emissions per CO2 unit time of the fuel consumption detected by the fuel sensor is ejected from the vehicle, and compares the calculated the instantaneous emissions and the threshold value, the comparison to evaluate the instantaneous emissions based on the results.
技術用語評価	B

表 4-1 重要技術用語抽出例 1

日本語の技術用語が一文の中で複数回現れた場合、その英語訳が一つに定まらない場合がある。Google 翻訳の翻訳結果で最初に現れる **feed section** だけを評価するか、中ほどの **power supply portion** も評価するかで評価値が異なることが考えられる。このような事例は統計ベース方式の Google 翻訳 と みんなの翻訳の翻訳結果で見られる。

公開番号	JP2014241702A
原文	給電部 20 を有し、走行自在に構成された移動パネル 1 への給電構造であって、前記移動パネルの上端部には、天井 7 側に設置され下向きに開口した開口部 5a を有した上レール 5 にガイドされる上ガイド部 3 が設けられ、前記上レールには、前記 給電部 に接続される電源コード 2 が前記移動パネルの移動に伴って追従するように収容され、前記開口部を塞ぐレールカバー部 50 が設けられていることを特徴とする。
重要技術用語	給電部
基準翻訳文	The power supply structure is an arrangement to supply electric power to a movable panel 1 which has a power supply unit 20 arranged movable. The movable panel has an upper guide section 3 provided at the upper part thereof guided by an upper rail 5 which is attached to a ceiling 7 and has an opening 5a opened downward. The upper rail stores a power cord 2 connected to the power supply unit so as to follow the movement of the movable panel. The upper rail has a rail cover section 50 which covers the opening.
重要技術用語	power supply unit
Google 翻訳	A feed section 20, a feed structure to the mobile panel 1 constructed running freely, and on the the upper part of the mobile panel, which had an opening 5a that opens downward is mounted on the ceiling 7 side the upper guide portion 3 to be guided disposed on the rail 5, the on rails, power cables 2 connected to the power supply portion is housed so as to follow with the movement of the moving panel, the opening that I am features a rail cover part 50 for closing are provided.
重要技術用語	feed section と power supply portion
技術用語評価	A

表 4-2 重要技術用語抽出例 2

4.3 評価対象機械翻訳文の抽出における課題

本調査で調査対象となる機械翻訳結果を日本語表示画面と英語表示画面から目視で対応する文を抽出したが、例えば、拒絶理由の場合、英語表示画面では段落が明示さないため、対応する文の抽出に当初の想定以上の時間と労力を要した（図 4-1）。この抽出で対応を誤った文を抽出した場合、抽出された文は本来抽出される文と似ているので、誤りが判明するのは機械翻訳文を精密に評価する時になる。誤りが見つかった場合は、対応する文の抽出まで戻って誤りを訂正して、再度評価することになり時間とコストを要した。このような場合、ファイル内の他の評価と一貫性がずれること、別の評価者に割り振られることがあり、評価の安定性にも影響する。

このような誤りは J-PlatPat、Google 翻訳と拒絶理由通知の機械翻訳文抽出で度々発生して機械評価のスケジュールに影響した。評価対象とする機械翻訳システムは 1 文単位に会話的に機械翻訳可能なシステムまたは対訳形式でデータが提供されることが望ましい。

・請求項 1-18 上記請求項に係る発明は、HVR1可変領域、HVR2可変領域及びigVR可変領域を包含するHCV-E2糖タンパク質において、igVR可変領域の少なくとも一部分が欠失している改変HCV-E2糖タンパク質に関するものである。 この点、本願明細書中には、HVR1可変領域、HVR2可変領域及びigVR可変領域を包含するHCV-E2糖タンパク質において、該領域における欠失対象部位を可動性リンカー配列で置換したこと、該置換により、CD81との結合のための立体構造が維持されたことが記載されているのみであり、単に欠失対象部位を欠失させたHCV-E2糖タンパク質変異体を作製し、CD81との結合活性を調べたことは具体的に全く記載されていない。 一般に、部分領域が欠失したタンパク質は、その立体構造が変動する蓋然性が高いことが出願時の技術常識であって、HVR1可変領域、HVR2可変領域及びigVR可変領域におけるアミノ酸の欠失が、HCV-E2糖タンパク質の立体構造に影響しないことが自明であったものとは認められない。 よって、上記請求項に記載された改変HCV-E2糖タンパク質の中には、立体構造が変動しCD81との結合活性が失われたものが含まれている蓋然性が高いものと認め、複数の遺伝子型タンパク質を得るに	- Claim: 1-18 In the HCV-E disaccharide protein which includes HVR1 variable region, HVR2 variable region, and an igVR variable region, the invention concerning the above-mentioned claim relates to the change HCV-E disaccharide protein in which at least 1 portion of the igVR variable region is carrying out deletion. In the HCV-E disaccharide protein which includes HVR1 variable region, HVR2 variable region, and an igVR variable region in this point and Description of this application, With having replaced the deletion subject part in the region by the mobile linker array, and the substitution, The HCV-E disaccharide protein variant to which it is [that it is only described that the spacial configuration for combination with CD81 was maintained and], and only carried out deletion of the deletion subject part is produced, and having investigated the avidity of CD81 is not described at all specifically. It is the technical common knowledge at the time of application that the protein, as for, subregion carried out deletion has generally the high probability of changing the spacial configuration, It is not accepted to be the obvious thing that the deletion of the amino acid in HVR1 variable region, HVR2 variable region, and an igVR variable region does not influence the spacial configuration of HCV-E disaccharide protein. Therefore, [in the change HCV E disaccharide protein described in the above-mentioned claim] The probability that that by which the spacial configuration was changed and the avidity of CD81 was lost is included is
---	--

図 4-1 評価対象機械翻訳文抽出例

項番 3「日英機械翻訳文の先行技術調査における有用性調査」では、検索システムにおいて、英語単語の語尾変化が検索式にヒットする/しないが影響することが分かった。今回と同様の調査を実施する場合、使用する英語検索式検証ツールには語尾変化を考慮した英語特有の表現が検索可能な検索システムによる検証が必要と考えられる。

Input search terms 

Query

digital-analog converter

Query Language:

Expansion Mode:

Precision  Recall



図 4-2 用語展開ツール例(PATENT SCOPE)

項番 2「日英機械翻訳の翻訳精度の調査」では公開特許公報以外にも拒絶理由通知書内の文も調査対象とした。拒絶理由通知書から評価対象の機械翻訳を抽出する際、下記に示すような、原文の「●2理由2・・・備考：」までの部分、それ以降の文が、ひとかたまりに機械翻訳されている1文の認識に失敗している例を発見した(J-PlatPatの下線部分)。このような見出し部分と本文とを適切に分け、見出し部分は翻訳メモリなど定形部分として翻訳するなどの工夫が必要と考えられる。

原文	J-PlatPat
●理由 2（同日出願）について	<u>the invention which relates to - Claim: 10 about Reason 2 (Double patenting) -- prior application 8 Remarks:</u> , [invention / (henceforth "the invention A") / concerning Claim 10 which quotes Claim 1 of this application] [the invention (henceforth "the invention B") concerning Claim 3 which quotes Claim 1 of the prior application 8 which Double patenting were carried out to this application, and was patented] Since the invention B is the same as that of the invention A when the invention A is made into a prior application and the invention B is made into later applications, and the invention A is the same as that of the invention B when the invention B is made into a prior application and the invention A is made into later applications, both correspond to "the same invention."
・請求項 10に係る発明について	
先願 8	
備考：	
本願の請求項 1 を引用する請求項 10 に係る発明（以下、「発明 A」という）と、本願と同日出願され特許された先願 8 の請求項 1 を引用する請求項 3 に係る発明（以下、「発明 B」という）とは、発明 A を先願とし発明 B を後願としたときに、発明 B は発明 A と同一であり、かつ、発明 B を先願とし発明 A を後願としたときに、発明 A は発明 B と同一であることから、両者は「同一の発明」に該当する。	

表 4-3 拒絶理由通知機械翻訳の不具合例

4.6 拒絶理由通知の引用文献の記述内容の引用

拒絶理由通知に引用文献の記述内容を括弧書きで嵌め込んで引用している場合、機械翻訳が嵌め込まれている部分を正しく認識できない場合がある。J-PlatPat、Google 翻訳、みんなの翻訳ともこの嵌め込まれた部分の認識に失敗している例を示す。何れも、本来は別の節として解析が必要な「センシング対象の～を備える受信装置」の構造理解が出来なかった。

出願番号	2013-124145
原文	引用文献1の図1、【0016】-【0019】には、「センシング対象の周波数帯域ごとに、前記周波数帯域で受信された信号が、自装置に中間周波数として設定されている第1の中間周波数の高調波の影響を受ける帯域の信号である場合には、前記第1の中間周波数から第2の中間周波数に切り替えて前記信号を受信し、高調波の影響を受ける帯域の信号ではない場合には、前記第1の中間周波数の切り替えを行わないように制御する制御部（クロックセクタ14）、を備える受信装置。」が記載されていると認められる。
基準翻訳文	Following description is noted in Fig.1 and [0016]-[0019] in the cited document 1: “Regarding a signal received in the said frequency band in every frequency band of a sensing object, if the signal is from the zone which is subject to the influence of the harmonics of the first intermediate frequency set on the self-device as an intermediate frequency, the device receives the signal switched from the above-mentioned first intermediate frequency to the second intermediate frequency. If it is not from the zone which is subject to the influence of the harmonics, the receiving device is provided with the control part (clock selector 14) controlled not to change the above-mentioned first intermediate frequency.”
J-PlatPat	[the Fig.1 of cited document 1, and [0016]-[0019]], [for every frequency band of a sensing object] [the signal received in the aforementioned frequency band] In being a signal of the zone which is subject to the influence of the harmonics of the first intermediate frequency set as the self-device as an intermediate frequency, In not being a signal of the zone which switches from the above-mentioned first intermediate frequency to a second intermediate frequency, receives the aforementioned signal, and is subject to the influence of harmonics, receiving set provided with the control part (clock selector 14) controlled not to change the above-mentioned first intermediate frequency. ” -- being described is admitted.
評価値	3
J-PlatPat の内容の伝達レベルの評価の根拠	(1) 「引用文献1の図1、【0016】-【0019】には、」が「が記載されていると認められる。」に係っているという基本構造が認識できなかったことで、本来は別の節として解析が必要な「センシング対象の～を備える受信装置」の構造理解が出来なかったようである。 (2) 「[(① {<センシング対象の周波数帯域ごとに>、前記周波数帯域で受信された信号が}、自装置に中間周波数として設定されている第1の中間周波数の高調波の影響を受ける帯域の信号である場合には)、(②前記第1の中間周波数から第2の中間周波数に切り替えて前記信号を受信し)、(③高調波の影響を受ける帯域の信号ではない場合には)、(④前記第1の中間周波数の切り替えを行わない) ように制御する制御部（クロックセクタ14）]、を備える受信装置。」において①+②&③+④が「のように制御する制御部」に係り、「備える」の目的語となり「受信装置」にかかる修飾節を構成しているという構文構造の解析に失敗したことで不完全な文生成となった。 (3) 節や句レベルの構成においては意味を伝えることが可能な訳出が多い。
Google 翻訳	Figure 1 of the cited reference 1, [0016] - [0019] The “for each frequency band of the sensing target, the signal received by said frequency band, a first intermediate that has been set as the intermediate frequency in the self-device If a band signal for receiving the harmonics influence of frequency, from said first intermediate frequency is switched to a second intermediate frequency to receive the signal, and if it is not the band of the signal affected by harmonics the, it is recognized that the control unit which controls not to perform the switching of the first intermediate frequency (clock selector 14), the receiving device comprises a.” is described.
評価値	2

Google 翻訳の 内容の伝達レ ベルの評価の 根拠	1) 「引用文献1の図1、【0016】-【0019】には、」が「が記載されていると認められる。」に係っているという基本構造が認識できなかったことで、本来は別の節として解析が必要な「センシング対象の～を備える受信装置」の構造理解が出来なかったようである。 2) 「引用文献1の図1、【0016】-【0019】」において「図1」と「【0016】-【0019】」が等位で結ばれ、そこに「引用文献1」が係っていることの解析が出来なかった。 3) 「[(① {<センシング対象の周波数帯域ごとに>、前記周波数帯域で受信された信号が}、自装置に中間周波数として設定されている第1の中間周波数の高調波の影響を受ける帯域の信号である場合には)、(②前記第1の中間周波数から第2の中間周波数に切り替えて前記信号を受信し)、(③高調波の影響を受ける帯域の信号ではない場合には)、(④前記第1の中間周波数の切り替えを行わない) ように制御する制御部(クロックセクタ14)]、を備える受信装置。」において①+②&③+④が「のように制御する制御部」に係り、備えるの目的語となり「受信装置」にかかる修飾節を構成しているという構文構造の解析に失敗したことで不完全な文生成となった。 4) ”a first intermediate that has been” と形容詞である”intermediate” を名詞のように扱い冠詞や関係代名詞節の付与が行われる非文法的な生成がある。
みんなの翻訳	FIG. 1 of Patent Document 1, the 【0016】1. - 【0019】1. and “in each frequency band of the sensing object, the signal received at the frequency band, a signal in a band affected by the intermediate frequency of the 1 harmonic is set as an intermediate frequency in the own device receives the signal from the intermediate frequency of the 1 switching to an intermediate frequency of 2 when there is no signal of the zone affected by the harmonics are controlled not to perform switching of the intermediate frequency of the 1 controller (clock selector 14), receiving apparatus.
評価値	2
みんなの翻訳 の評価の根拠	1) 「引用文献1の図1、【0016】-【0019】には、」が「が記載されていると認められる。」に係っているという基本構造が認識できなかったことで、本来は別の節として解析が必要な「センシング対象の～を備える受信装置」の構造理解が出来なかったようである。 2) 「引用文献1の図1、【0016】-【0019】」において「図1」と「【0016】-【0019】」が等位で結ばれ、そこに「引用文献1」が係っていることの解析が出来なかった。 3) 「[(① {<センシング対象の周波数帯域ごとに>、前記周波数帯域で受信された信号が}、自装置に中間周波数として設定されている第1の中間周波数の高調波の影響を受ける帯域の信号である場合には)、(②前記第1の中間周波数から第2の中間周波数に切り替えて前記信号を受信し)、(③高調波の影響を受ける帯域の信号ではない場合には)、(④前記第1の中間周波数の切り替えを行わない) ように制御する制御部(クロックセクタ14)]、を備える受信装置。」において①+②&③+④が「のように制御する制御部」に係り、備えるの目的語となり「受信装置」にかかる修飾節を構成しているという構文構造の解析に失敗したことで不完全な文生成となった。 4) 「第1の中間周波数の高調波」において 訳出の”the intermediate frequency of the 1 harmonic” では係りが逆になる日本語と同じ前置修飾となっている。また、序数は数字ではなく first としなければならない。

表 4-4 拒絶理由通知機械翻訳の引用箇所の不具合例

4.7 品質評価手順の改善案

項番 4.1「内容の伝達レベルの評価作業に関する問題点・課題」から項番 4.6「拒絶理由通知の引用文献の記述内容の引用」までに示した問題点・課題等を踏まえ、特許庁の「品質評価手順」について改善案を以下に示す。

4.7.1 調査対象文の機械翻訳文の取得方法

項番 4.3「評価対象機械翻訳文の抽出における課題」と 4.5「拒絶理由通知書の見出し部分の翻訳不備」で示したように、調査対象文に対応する機械翻訳文を抽出する作業に「日本語原文に対応する機械翻訳文の抽出が難しい」という課題が存在する。この抽出誤りに気が付くのは機械翻訳の評価に入ってからになるため戻り作業が大きく、再評価は評価と別のタイミングになり、時には別の評価者に割り振られてしまうため、一貫性のある評価を妨げることになる。

例えば、J-PlatPat であれば翻訳前後の原文と機械翻訳文を対応付けることで対のデータが作成可能である。そのような機械翻訳文が存在しているのであれば、そのデータを利用することで機械翻訳文抽出作業の効率化が図ることができる。また、espacenet の翻訳結果についても、espacenet の公報全文の英訳から原文と対応する機械翻訳文を見つけるより、NICT のみんなの自動翻訳の様に原文をテキスト翻訳ボックスに貼り付けて機械翻訳文を得る方が能率と精度良く機械翻訳文を入手できる。

原文と機械翻訳文の対応精度が正しくすることは評価基準以前の問題であるが、実際には評価精度そのものに大きく影響する。

4.7.2 1 文献からの調査対象文の抽出数について

本調査では、「調査対象とした文献から 3 文の調査対象文を抽出する」仕様となっていたが、例えば、一つの文献から 3 文を抽出する場合、2 文までは抽出条件を満たすが、最後の 3 文目の条件が合わずに、その文献が採用できなかった場合があった。また、一つの文献から抽出すると似たような用語から成る文が集まってしまう。バリエーションに富んだ文を調査対象とする観点から、1 文献から 1 文を抽出した方が好ましい。また、何らかの理由で対象文献を入れ替える必要がある場合、影響範囲が 1 文で済む利点も考えられる。

4.7.3 自動評価の活用

項番 2.2.3「人手による評価と自動評価の相関関係」で「内容の伝達レベル」評価結果と自動評価手法「RIBES」の評価結果との間に、若干の相関関係がある結果を得た。この結果から人手による評価に自動評価を活用する方法を検討した。

具体的には、「内容の伝達レベル」の評価を行う前に、自動評価手法「RIBES」の評価結果を取得し、評価結果の低い文を優先して「内容の伝達レベル」調査を行うなどの活用

が考えられる。但し、「内容の伝達レベル」と「RIBES」の相関関係は本調査で行った日英機械翻訳についての結果であるため、他の言語対で評価を行う場合は、別途事前調査が必要となる。

また、利用している機械翻訳エンジンの改善と副作用を効率的に測る目的で、新旧のシステムで自動評価値が大きく変化した順に機械翻訳結果を人手評価することが考えられる。例えば、自動評価値が大きく上がった場合は機械翻訳精度が向上した可能性が、評価値が大きく下がった場合は副作用による劣化の可能性がある。

このように自動評価で目的に応じて評価対象文を選定することで、全ての評価対象文を平等に人手評価するより、早く低コストで人手評価できる可能性がある。

5. 付録

5.1 WIPO 技術分野一覧

項番	セクター名	35 分類との対応
1	Electrical engineering (電気工学)	1～ 8
2	Instruments (機器)	9～13
3	Chemistry (化学)	14～24
4	Mechanical engineering (機械工学)	25～32
5	Other fields (その他)	33～35

IPC8-Technology Concordance(セクター単位) (再掲)

分野番号	分野名	分野番号	分野名
1	Electrical machinery, apparatus, energy	19	Basic materials chemistry
2	Audio-visual technology	20	Materials, metallurgy
3	Telecommunications	21	Surface technology, coating
4	Digital communication	22	Micro-structural and nano-technology
5	Basic communication processes	23	Chemical engineering
6	Computer technology	24	Environmental technology
7	IT methods for management	25	Handling
8	Semiconductors	26	Machine tools
9	Optics	27	Engines, pumps, turbines
10	Measurement	28	Textile and paper machines
11	Analysis of biological materials	29	Other special machines
12	Control	30	Thermal processes and apparatus
13	Medical technology	31	Mechanical elements
14	Organic fine chemistry	32	Transport
15	Biotechnology	33	Furniture, games
16	Pharmaceuticals	34	Other consumer goods
17	Macromolecular chemistry, polymers	35	Civil engineering
18	Food chemistry		

5.2 自動評価手法の概要

機械翻訳の評価手法として古くから人手による評価が行われ、現在も行われている。しかしながら、人手による評価はコストがかかるため、コンピュータを用いた自動評価ができれば望ましい。そのような動機から自動評価が考えられ、これまでに多くの自動評価基準が提案されている。自動評価では、まず試験文(原文)に対して基準翻訳文と呼ばれる、いわば理想的な訳文を用意しておく。その後、原文を機械翻訳し、機械翻訳文と基準翻訳文とがどれほど近いかで翻訳精度を評価するものである。基準翻訳文に近い機械翻訳文ほど翻訳精度が高いと判断する。このときの「近さ」をどのようにして測るかで様々な自動評価基準が考えられる¹⁰。自動評価については[1]で既に解説されているが¹¹、ここでは、今回使用した3種類の自動評価基準である BLEU、NIST、RIBES について、具体例を交えて説明する。

5.2.1 BLEU

BLEU(Bilingual Evaluation Understudy)は機械翻訳の自動評価基準として最初に提案された基準であり[2]、現在でも多く使われている。BLEU は以下のようにして求められる。基準翻訳文と機械翻訳文の双方を単語分割し¹²、得られる単語 n グラム(連続する n 個の単語) $w_1, w_2, \dots, w_n$ に対して、基準翻訳文と機械翻訳文中の頻度を求める。

$C^{tst}(w_1, w_2, \dots, w_n)$ = 機械翻訳文中の $w_1, w_2, \dots, w_n$ の頻度

$C^{ref}(w_1, w_2, \dots, w_n)$ = 基準翻訳文中の $w_1, w_2, \dots, w_n$ の頻度

そして、 $C_{clip}^{tst}(w_1, w_2, \dots, w_n)$ を次のようにして求める。

$$C_{clip}^{tst}(w_1, w_2, \dots, w_n) = \min(C^{tst}(w_1, w_2, \dots, w_n), C^{ref}(w_1, w_2, \dots, w_n))$$

つまり、機械翻訳文中の頻度と基準翻訳文中の頻度のうちの小さい方の値である。具体例で説明する。

[例文 1]

原文 コネクタ本体 14 が基板 12 上に実装される

基準翻訳文 The connector main body 14 is mounted on the substrate 12

機械翻訳文 The connector body 14 is mounted on the substrate body 12

とするとき、 $n=1$ として $w_1 = \text{"body"}$ の場合を考える。このとき、機械翻訳文中の "body" の頻度は 2 であり、基準翻訳文中の "body" の頻度は 1 である。そこで、 $C^{tst}(\text{body}) = 2$ 、

¹⁰ 原文を翻訳する場合、理想的な訳文は一つとは限らない。そこで、基準翻訳文を複数用意して近さを測る方法もあるが、特許翻訳の場合、単一の基準翻訳文を用いることが多く、今回の評価でもそのようにしている。そこで、本文では単一の基準翻訳文の場合についてのみ説明する。また、1 文単位で評価するのではなく複数の文からなるドキュメント単位で評価する場合があるが、ここでは文単位の評価について考える。

¹¹ 第 4 章 自動評価、4.1 自動評価手法を参照。

¹² ここでは、大文字と小文字を区別せず、 "The" と "the" は同一の単語とみなす。また、ピリオドやコンマなどはその前の単語と分離して単独の単語として扱う。

$C^{ref}(body) = 1$ 、 $C_{clip}^{tst}(body) = 1$ となる。

次に、機械翻訳文に含まれるすべての n グラム G_n を用いて n グラムの適合率 p_n を以下の式で求める。

$$p_n = \frac{\sum_{w_1, w_2, \dots, w_n \in G_n} C_{clip}^{tst}(w_1, w_2, \dots, w_n)}{\sum_{w_1, w_2, \dots, w_n \in G_n} C^{tst}(w_1, w_2, \dots, w_n)}$$

例文 1 の場合、 $G_1 = \{the, connector, body, 14, is, mounted, on, substrate, 12\}$ であり、

$$p_1 = \frac{2 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1}{2 + 1 + 2 + 1 + 1 + 1 + 1 + 1 + 1} = \frac{10}{11} = 0.909$$

となる。 p_n の相乗平均をもとに BLEU は以下のように計算される。

$$BLEU = BP \cdot \sqrt[N]{p_1 p_2 \dots p_N}$$

ここで、 N は考慮する最大のグラム数であり、通常は 4 が用いられる。また、 BP は brevity penalty と呼ばれる補正項であり、機械翻訳文が短い場合に BLEU の値が大きくなってしまう不具合を修正するもので

$$BP = \begin{cases} 1, & c > r \\ e^{(1-r/c)}, & c \leq r \end{cases}$$

で定義される¹³。ここで c は機械翻訳文の単語数であり、 r は基準翻訳文の単語数である。つまり、 c が r より少ないときにペナルティを課す項である。たとえば、例文 1 の場合、機械翻訳文が”body 14 is mounted”という断片的な訳文でも $N=4$ としたときには $p_1 = p_2 = p_3 = p_4 = 1$ となり $\sqrt[4]{p_1 p_2 \dots p_N} = 1$ となってしまうが、 $r/c = 11/4 = 2.75$ となり、 $BP = 0.174$ であるから $BLEU = 0.174$ と小さな値となる。

$0 \leq p_n \leq 1$ が成り立ち、 $0 < BP \leq 1$ であるから、BLEU は 0 と 1 の間の数値をとり、大きいほど機械翻訳文が基準翻訳文に近いことになる。

例文 1 の場合、 $p_1 = 0.9091$, $p_2 = 0.7$, $p_3 = 0.5556$, $p_4 = 0.5$ となり、 $\sqrt[4]{p_1 p_2 p_3 p_4} = 0.6484$ である。さらに $BP = 1$ であるため $BLEU=0.6484$ となる。

5.2.2 Smoothed BLEU

BLEU の問題点のうちの一つは、 p_n の相乗平均を用いているため、 $p_n = 0$ となる n が一つでもあると $BLEU=0$ となってしまうことである。このため、オリジナルの BLEU では文単位の評価に用いるのではなく、ドキュメント単位の評価に用いるべきものとしている。ドキュメントにおいては 4 グラムのような比較的大きな n に対しても $p_n = 0$ となることは考えにくく、その場合 $BLEU=0$ となることもないからである。しかし BLEU を文単位の評価に用いたいという要望は多く、上記の問題に対処する方法がいくつか提案されている[3]。ここでは NIST からフリーの評価ツール mteval-v13a.pl として提供されている方法について説明する。 n グラムの適合率を以下のように修正する。

¹³ c/r に対する BP の値は後述する図 1 に示されている。

$$\tilde{p}_n = f(x) = \begin{cases} \frac{\sum_{w_1, w_2, \dots, w_n \in G_n} C_{clip}^{tst}(w_1, w_2, \dots, w_n)}{\sum_{w_1, w_2, \dots, w_n \in G_n} C^{tst}(w_1, w_2, \dots, w_n)}, & p_n \neq 0 \\ \frac{1/2^{(n-k+1)}}{\sum_{w_1, w_2, \dots, w_n \in G_n} C^{tst}(w_1, w_2, \dots, w_n)}, & n \geq k \text{ where } p_{k-1} \neq 0 \text{ and } p_k = 0 \end{cases}$$

例えば、 $p_4 = 0$ で $p_3 \neq 0$ の場合、 $\tilde{p}_4$ の分子は、 $1/2^{(4-4+1)} = 1/2$ となる。 $\tilde{p}_n$ を用いて

$$BLEU^{\sim} = BP * \sqrt[N]{\tilde{p}_1 \tilde{p}_2 \dots \tilde{p}_N}$$

として smoothed BLEU が求められる。

具体例を示す。以下の例文 2 を考える。なお、以下で用いる例文の原文と基準翻訳文はすべて例文 1 と同じであるので記述を省略する。

[例文 2]

機械翻訳文 Connector main part 14 is mounted in substrate 12

機械翻訳文に含まれる 4 グラムは{Connector main part 14, main part 14 is, part 14 is mounted, 14 is mounted in, is mounted in substrate, mounted in substrate 12}であるが、これらはいずれも基準翻訳文には含まれていない。したがって $p_4 = 0$ となり、オリジナルの BLEU は 0 となってしまう。一方、 $\tilde{p}_4 = 1/(2 \times 6) = 0.0833$ となり、 $BLEU^{\sim} = 0.2089$ となる。

5.2.3 NIST

NIST (National Institute of Standards and Technology)¹⁴は基本的に BLEU の考え方を継承し、改良を加えた評価基準であり、以下の式で定義される[4]。

$$NIST = BP_{NIST} \sum_{n=1}^N q_n$$

ここで q_n は

$$q_n = \frac{\sum_{w_1, w_2, \dots, w_n \in G_n} \text{Info}(w_1, w_2, \dots, w_n)}{\sum_{w_1, w_2, \dots, w_n \in G_n} 1}$$

で与えられる。N は通常 5 が用いられる。NIST は 0 以上であるが、BLEU と異なり 1 以上の値をとることもある。定義式において BLEU と異なる点は 3 点ある。

- (1) BLEU は相乗平均を用いていたが NIST では和を用いている。
- (2) brevity penalty の計算方法が異なる。
- (3) q_n の式の分子の関数形が BLEU では $C_{clip}^{tst}(w_1, w_2, \dots, w_n)$ であったが NIST では $\text{Info}(w_1, w_2, \dots, w_n)$ となっている。分母も BLEU では機械翻訳文に含まれる n グラムの延べ度数であったが、NIST では異なり度数となっている。

(1)によって $q_n = 0$ となる n があっても NIST=0 となるとは限らない。つまり smoothing を必ずしも必要としない。(2)に関しては

$$BP_{NIST} = \exp\{\beta \log^2 \left[\min \left(\frac{c}{r}, 1 \right) \right]\}$$

である。 β は $c/r=2/3$ のときに $BP_{NIST} = 0.5$ となるように定められる。 c/r に対する BP と BP_{NIST} の値の変化を図 1 に示す。 $c/r > 0.761$ の時に $BP_{NIST} \geq BP$ となり、 $c/r < 0.761$ の時には逆になる。

¹⁴ NIST は組織の名前であるが、この自動評価基準は、NIST が主催している評価型 workshop で用いられており、NIST の名で呼ばれている。

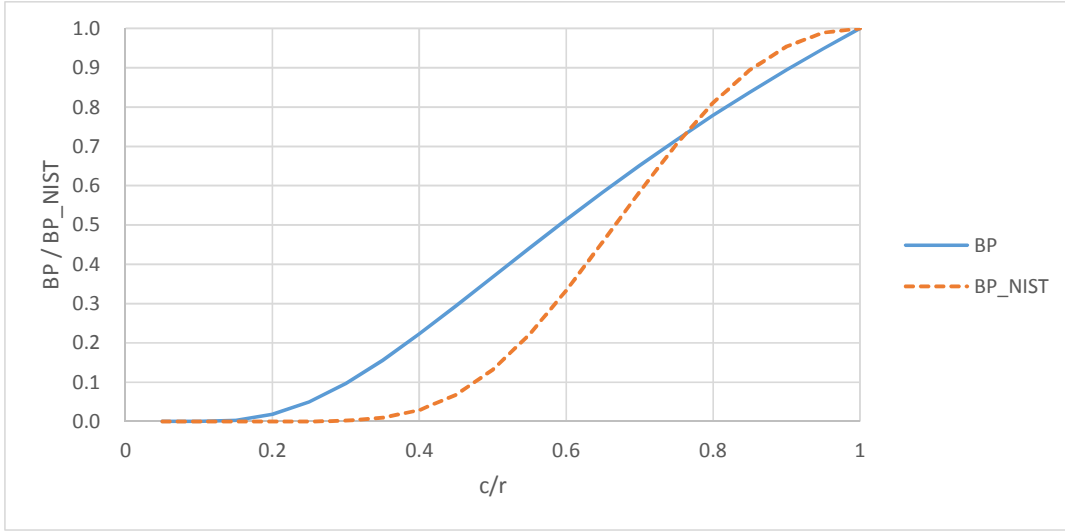


図 5-1 基準翻訳文の単語数(r)と機械翻訳文の単語数(c)の比に対する brevity penalty

(4)に示す Info は以下のように定義され、基準翻訳文において、 $w_1, w_2, \dots, w_{n-1}$ が持つ情報量から $w_1, w_2, \dots, w_n$ が持つ情報量への増加分を表している。ここで $C^{ref}(w_1, w_2, \dots, w_n)$ は前述した基準翻訳文に含まれる $w_1, w_2, \dots, w_n$ の頻度である。

$$\text{Info}(w_1, w_2, \dots, w_n) = \begin{cases} 0, & C^{ref}(w_1, w_2, \dots, w_n) = 0 \\ \log_2\left(\frac{C^{ref}(w_1, w_2, \dots, w_{n-1})}{C^{ref}(w_1, w_2, \dots, w_n)}\right), & \text{else} \end{cases}$$

ただし $n=1$ の場合は、 $C^{ref}(w_1, w_2, \dots, w_{n-1})$ を基準翻訳文に含まれる全単語数とする。

例文 1 において Info は以下ようになる。

$$\text{Info}(\text{the}, \text{substrate}) = \log_2\left(\frac{C^{ref}(\text{the})}{C^{ref}(\text{the}, \text{substrate})}\right) = \log_2\left(\frac{2}{1}\right) = 1$$

$$\text{Info}(\text{mounted}, \text{on}) = \log_2\left(\frac{C^{ref}(\text{mounted})}{C^{ref}(\text{mounted}, \text{on})}\right) = \log_2\left(\frac{1}{1}\right) = 0$$

$$\text{Info}(\text{the}) = \log_2\left(\frac{11}{C^{ref}(\text{the})}\right) = \log_2\left(\frac{11}{2}\right) = 2.46$$

$$\text{Info}(\text{mounted}) = \log_2\left(\frac{11}{C^{ref}(\text{mounted})}\right) = \log_2\left(\frac{11}{1}\right) = 3.46$$

となる。例文 1 に対して、 $q_1 = 2.9631$, $q_2 = 0.2$, $q_3 = 0$, $q_4 = 0$ となり、さらに $BP_{NIST} = 1$ となるため $NIST=3.1631$ である。

5.2.4 RIBES

BLEU と NIST のもう一つの問題点は、n グラムという局所的な情報のみを用いているため、語順の大幅な入れ替えがある日英翻訳などの評価に用いた場合、評価の信頼性が劣る点である。実際 NTCIR-9 の特許翻訳タスク、日英サブタスクでは、BLEU や NIST の値と人手評価結果(Adequacy)との相関係数が負になっている[5]¹⁵。そこで大域的な語順の類似性も考慮した自動評価基準として RIBES が提案されている[6]。RIBES では基準翻訳文と機械翻訳文との語順の類似性を Kendall の順位相関係数を用いて求めている。Kendall の順位相関係数は次のようにして計算される。まず機械翻訳文と基準翻訳文との間に単語対応を求め、2 個の単語対応の組み合わせについて単語の出現順序が機械翻訳文と基準翻訳文との間で同順か逆順かの場合数を計数する。

具体例で説明する。

[例文 3]

機械翻訳文 On the substrate 12 , the connector main body 14 is mounted
の場合を考えると、単語対応が図 2 のように得られる。

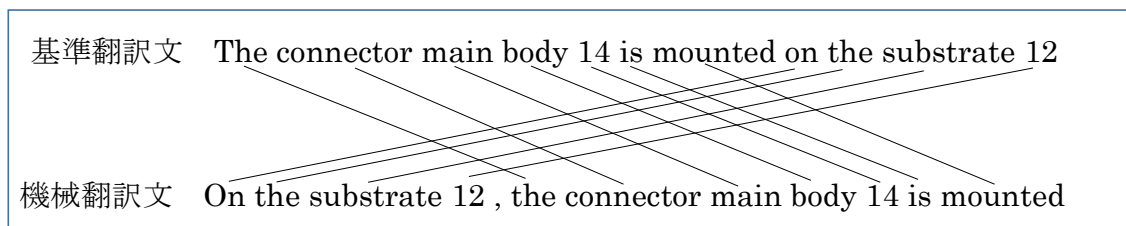


図 5-2 単語対応例

このとき 2 個の単語対応の組み合わせに対して同順(S)と逆順(I)が図 3 のように求まる。例えば、on と substrate の組み合わせの場合、基準翻訳文での出現位置は、おののおの 8 と 10 であり、機械翻訳文での出現位置は 1 と 3 である。つまり基準翻訳文と機械翻訳文のいずれでも on より substrate のほうが後に出現しており、同順となる。一方、on と connector では、基準翻訳文での出現位置が、おののおの 8 と 2 であるのに対して、機械翻訳文での出現位置は 1 と 6 である。そこで逆順となる。

¹⁵文献[5]の Table 19 (page 573)を参照。Pearson の相関係数が BLEU で -0.241 、NIST で -0.286 である。

	the	substrate	12	the	connector	main	body	14	is	mounted
on	S	S	S	I	I	I	I	I	I	I
the		S	S	I	I	I	I	I	I	I
substrate			S	I	I	I	I	I	I	I
12				I	I	I	I	I	I	I
the					S	S	S	S	S	S
connector						S	S	S	S	S
main							S	S	S	S
body								S	S	S
14									S	S
is										S

図 5-3 例文 3 の単語対応の同順(S)と逆順(I)

同順の個数を C_S とし、逆順の個数を C_I とする。また、単語対応の個数を n とすると、2 個の単語対応の組み合わせの総数は $n(n-1)/2$ である。このとき Kendall の順位相関係数(τ)は

$$\tau = \frac{C_S - C_I}{\frac{n(n-1)}{2}}$$

で与えられる。2 個の単語対応の組み合わせについて、すべてが同順であれば $\tau = 1$ となり、全てが逆順になれば $\tau = -1$ となる。さらに τ の範囲を $[0,1]$ に正規化した NKT (Normalized Kendall's τ)を

$$\text{NKT} = \frac{\tau + 1}{2}$$

と定義する。

例文 3 の場合は

$$\tau = \frac{27 - 28}{\frac{11(11-1)}{2}} = -0.018$$

$$\text{NKT} = \frac{-0.018 + 1}{2} = 0.491$$

となる。

τ や NKT には次の性質があり、このままでは自動評価基準として不適切である。

- (1) 単語対応の個数が少なく、2 個の単語対応の組み合わせに同順が多い場合に高い値となる。
- (2) 機械翻訳文が短く、2 個の単語対応の組み合わせに同順が多い場合に高い値となる。

例えば、

[例文 4]

機械翻訳文 A connected major part 15 was inserted in the substrate 12

とすると、単語対応は(the, substrate, 12)のみであり、2 個の単語対応の組み合わせは全て同順であるので NKT=1 となる。また

[例文 5]

機械翻訳文 is mounted

に対しても、NKT=1 となる。例文 4 は(1)の結果であり、例文 5 は(2)の結果である。

RIBES では、上記問題点の(1)に対応するために以下に述べるペナルティ P を用いている。また、(2)に対応するために BLEU の brevity penalty と同じ BP を用いている。ここで、P は以下の式で定義される。

$$P = \frac{n}{c}$$

n は単語対応の個数、c は機械翻訳文の単語数である。最終的に RIBES は

$$\text{RIBES} = \text{NKT} \cdot P^\alpha \cdot BP^\beta$$

として定義される。ここで $\alpha \geq 0$ と $\beta \geq 0$ はパラメータであり、 $\alpha = 0.25$ 、 $\beta = 0.1$ が初期設定されている。RIBES は 0 以上 1 以下の値をとり、大きいほど基準翻訳文と機械翻訳文が近いと評価される。

[例文 6]

機械翻訳文 The substrate 12 is mounted on the connector main body 14

とすると、例文 6 は基準翻訳文とまったく異なる意味を持つが、局所的な n グラムの範囲では基準翻訳文と類似している。その結果、BLEU = 0.375 となる。一方、大域的語順を考慮した RIBES では、RIBES = 0.273 と BLEU に比較して小さい値となる。

実際、前述した NTCIR-9 のデータ[5]では、BLEU や NIST の値と人手評価結果(Adequacy)との相関係数が負になっていたが、RIBES の値と Adequacy との相関係数は正となっている¹⁶。

参考文献

- [1] 特許庁：特許審査関連情報の機械翻訳による英語提供に対する精度評価に係る調査報告書、平成 23 年 2 月。
- [2] Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu：BLEU：a Method for Automatic Evaluation of Machine Translation, *Proceedings of the 40th Annual*

¹⁶ 文献[5]の Table 19 (page 573)を参照。Pearson の相関係数が 0.579 である。

- Meeting of the Association for Computational Linguistics*, pp. 311–318, 2002.
- [3] Boxing Chen and Colin Cherry : A Systematic Comparison of Smoothing Techniques for Sentence-Level BLEU, *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pp. 362–367, 2014.
- [4] George Doddington: Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics, *Proceedings of the second international conference on Human Language Technology Research*, pp.138–145, 2002.
- [5] Isao Goto, Bin Lu, Ka Po Chow, Eiichiro Sumita and Benjamin K. Tsou: Overview of the Patent Machine Translation Task at the NTCIR-9 Workshop, *Proceedings of NTCIR-9 Workshop Meeting*, Tokyo, Japan, pp.559–578, 2011.
- [6] 平尾努、磯崎秀樹、須藤克仁、Duh Kevin、塚田元、永田昌明：語順の相関に基づく機械翻訳の自動評価法、*自然言語処理*、Vol.21, No.3, pp.421–444, 2014.
- [7] Hideki Isozaki and Natsume Kouchi: Dependency Analysis of Scrambled References for Better Evaluation of Japanese Translation, *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pp. 450–456, 2015.

5.3 評価者間の評価の一致度合いを調べる指標

評価者が2名(R1とR2とする)の場合を考える。2名が評価した文数を N とする。評価カテゴリーの総数を K とする。相対評価の場合は $K=3$ である(システムAがシステムBより良い、悪い、同等の3カテゴリー)。重要技術用語の翻訳精度に関する絶対評価の場合は $K=4$ である(A, B, C, D)。ここで、 $1 \leq i, j \leq K$ に対してR1が i と評価し、R2が j と評価した文数を n_{ij} とし、 $n_{.j} = \sum_{i=1}^K n_{ij}$ 、 $n_{i.} = \sum_{j=1}^K n_{ij}$ 、 $n_{..} = \sum_{i=1}^K \sum_{j=1}^K n_{ij}$ とする。 $n_{..} = N$ である。2名の評価者の評価が一致している文の割合は

$$p_o = \sum_{i=1}^K \frac{n_{ii}}{N}$$

である。 p_o 自体を一致度合いの指標として用いることもできるが、両者の評価が偶然に一致する割合との比として一致度合いの指標を定義するのが通例である。偶然に一致する割合として2種類が提案されている。Cohen's kappa の場合には、

$$p_e = \sum_{i=1}^K \left(\frac{n_{i.}}{N} \cdot \frac{n_{.i}}{N} \right)$$

で定義し Scott's Pi の場合には

$$p'_e = \sum_{i=1}^K \left(\frac{n_{i.} + n_{.i}}{2N} \cdot \frac{n_{i.} + n_{.i}}{2N} \right)$$

で定義する。そして、Cohen's kappa は

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

で与えられ、Scott's Pi は

$$\pi = \frac{p_o - p'_e}{1 - p'_e}$$

で与えられる。 κ の範囲として

$$\frac{-p_e}{1 - p_e} \leq \kappa \leq 1$$

が成り立ち、 π も同様である。 $p_o = 1$ の場合は $\kappa = 1$ であり、 $p_o = p_e$ の場合は $\kappa = 0$ である。 $p_o < p_e$ の場合は $\kappa < 0$ となる。 κ の解釈については以下が提案されている[1]。

κ	interpretation
0.0-0.2	slight
0.2-0.4	fair
0.4-0.6	moderate
0.6-0.8	substantial
0.8-1.0	almost perfect

参考文献

[1] Landis and Koch. 1977. The measurement of observer agreement for categorical data, *Biometrics*, Vol. 33, pp.159-174.