

平成 25 年度
特許文献機械翻訳の
品質評価手法に関する調査

調 査 報 告 書

平成 26 年 2 月

一般財団法人 **日本特許情報機構**

目 次

第1章 調査の目的と概要	11
1.1. 調査の目的	11
1.2. 調査内容	12
1.3. 調査方法の概要	12
1.3.1. 文献調査	12
1.3.2. 英日機械翻訳の現状分析	13
1.3.3. 新設機械翻訳品質評価手法の作成	13
1.3.4. 委員会による検討	13
1.3.5. 特許庁機械翻訳品質評価手法の作成	14
1.3.6. 英日機械翻訳に関するテストセットの作成	14
1.3.7. 英日機械翻訳の実施	14
1.3.8. 特許庁機械翻訳品質評価手法の実践	14
1.3.9. 品質評価結果の分析	14
1.3.10. 報告書の作成	14
1.4. 調査結果の概要	16
1.4.1. 文献調査	16
1.4.2. 英日機械翻訳の現状分析	18
1.4.3. 「特許庁機械翻訳品質評価手法」の作成	22
1.4.4. 英日機械翻訳に関するテストセットの作成	26
1.4.5. 特許庁機械翻訳品質評価手法の実践	28
1.4.6. 「特許庁手法」の実用性と今後の課題	31
第2章 文献調査	37
2.1. 翻訳品質評価手法の内容	37
2.1.1. 機械翻訳品質評価に関する主要な先行研究	37
2.1.2. NTCIR-9/NTCIR-10での人手評価	48
2.1.3. 過去の特許庁調査における翻訳品質評価手法の内容	57
2.1.4. 本調査に有用な知見	62
2.2. 英日機械翻訳の内容	65
2.2.1. 市販ソフトウェア（国内）	65
2.2.2. 市販ソフトウェア（海外）	67
2.2.3. ウェブ上で利用可能な無料翻訳サービス	68
2.2.4. 研究機関の英日機械翻訳エンジン	69
2.2.5. 本調査に有用な知見	70
第3章 英日機械翻訳の現状分析	75
3.1. 英日機械翻訳における重要技術用語の翻訳精度の評価	75

3.1.1. 英日機械翻訳エンジン	75
3.1.2. サンプル英語特許文献	75
3.1.3. 重要技術用語の選定	76
3.1.4. 用語の翻訳精度の評価基準	77
3.1.5. 重要技術用語の翻訳精度評価サンプル	78
3.1.6. 評価結果の集計方法	78
3.1.7. 評価結果	79
3.2. 英日機械翻訳の特有課題・留意点及び対応策	85
3.2.1. 英日機械翻訳の特有課題・留意点	85
3.2.2. 英日機械翻訳特有課題への対処策	94
第4章 特許庁機械翻訳品質評価手法の作成	97
4.1. 特許庁機械翻訳品質評価手法の前提条件	97
4.1.1. 機械翻訳の対象	97
4.1.2. 機械翻訳結果の用途	97
4.1.3. 品質評価の目的・利用方法	98
4.1.4. 評価者	99
4.1.5. 評価作業のリソース制限	99
4.2. 特許庁機械翻訳品質評価手法の構成	100
4.3. 特許庁機械翻訳品質評価手法の内容（1）：「人手による品質評価の実施」部分 ..	102
4.3.1. 評価の観点及び基準：（1）内容の伝達レベルの評価	102
4.3.2. 評価の観点及び基準：（2）重要技術用語の翻訳精度の評価	109
4.3.3. 評価の観点及び基準：（3）流暢さの評価	114
4.3.4. 評価の観点及び基準：（4）チェックリストに基づく評価①「基本方針」	119
4.3.5. 評価の観点及び基準：（4）チェックリストに基づく評価②【必須項目】	123
4.3.6. 評価の観点及び基準：（4）チェックリストに基づく評価③【任意項目】	140
4.3.7. 評価の観点及び基準：（4）チェックリストに基づく評価④【翻訳言語別任意項目】	155
4.3.8. 相対評価方法	157
4.3.9. 絶対評価方法	162
4.3.10. 主要項目と補助項目	166
4.4. 特許庁機械翻訳品質評価手法の内容（2）：「事前準備」部分	168
4.4.1. テストセットの準備	168
4.4.2. 評価作業用データの準備	173
4.4.3. 評価支援ツールの準備	174
第5章 英日機械翻訳に関するテストセットの作成	177
5.1. テストセットの要件	177
5.1.1. 「特許庁手法」のテストセット定義	177
5.1.2. 英日テストセットの要件	178

5.2. 英日テストセットの作成手順	179
5.2.1. 英日文アライメントデータの作成と英日テストセット候補データの抽出	179
5.2.2. 英日テストセットにおける文の比率の確定	179
5.2.3. 文の選定と目視チェック	184
5.2.4. チェックリスト項目検証用フラグのセット	186
5.2.5. 重要技術用語の特定	201
5.3. 英日テストセットのフォーマット	204
5.3.1. 英日テストセットのファイル形式とデータ構成	204
5.3.2. データ行の構成	205
第6章 英日機械翻訳の実施	209
6.1. 評価対象文の選定	209
6.1.1. 記載箇所	209
6.1.2. 技術分野（IPC）	210
6.1.3. 原文の長さ	210
6.1.4. その他の構成比率	211
6.1.5. 検証用フラグの確認	211
6.2 英日機械翻訳の実施	213
6.2.1. 英日機械翻訳エンジン	213
6.2.2. 文分割に対する処理	213
6.3. 品質評価対象データの作成	215
6.3.1. 英日テストセットからの文の抽出	215
6.3.2. 機械翻訳結果のセット	215
6.3.3. 品質評価作業用ファイルの作成	215
第7章 特許庁機械翻訳品質評価手法の実践	223
7.1. 「特許庁手法」の実践内容	223
7.1.1. 評価観点及び対象文数	223
7.1.2. 評価者	223
7.1.3. 評価観点及び評価基準の周知	224
7.1.4. 人手評価補助ツール	225
7.1.5. 評価実践にあたっての留意点	226
7.2. 品質評価結果データ	227
7.2.1. 品質評価結果の統合	227
7.2.2. 分析用情報の追加	227
7.2.3. 「品質評価結果データ」のレイアウト	227
第8章 品質評価結果の分析	231
8.1. 「品質評価結果データ」に関する注意事項	231
8.1.1. 機械翻訳結果の作成条件	231

8.1.2. データ件数.....	231
8.2. 評価結果の分析と考察：「内容の伝達レベルの評価」	232
8.2.1. 評価基準	232
8.2.2. 評価スコア	233
8.2.3. 評価者別の評価値.....	233
8.2.4. 符号検定による検証	235
8.2.5. 評価対象文数の妥当性の検証	237
8.3. 評価結果の分析と考察：「重要技術用語の翻訳精度の評価」	238
8.3.1. 評価基準	238
8.3.2. 評価スコア	238
8.3.3. 符号検定による検証	239
8.4. 評価結果の分析と考察：「流暢さの評価」	240
8.4.1. 評価基準	240
8.4.2. 評価スコア	240
8.4.3. 符号検定による検証	241
8.4.4. 評価対象文数の妥当性の検証	242
8.5. 評価結果の分析と考察：「チェックリストに基づく評価」	243
8.5.1. 評価基準	243
8.5.2. 【必須項目】カウント集計結果.....	243
8.5.3. 【任意項目】カウント集計結果.....	245
8.5.4. 【翻訳言語別任意項目 [英日]】カウント集計結果.....	246
8.6. 絶対評価スコアのマッチング基準の妥当性検証.....	247
8.6.1. 妥当性検証の方法.....	248
8.6.2. 集計結果	248
8.6.3. マッチング基準の見直しの観点.....	249
8.6.4. 内容伝達スコアの閾値の検討	250
8.6.5. 技術用語適訳率の閾値の検討	250
8.6.6. 新マッチング基準とその検証	251
8.7. 「特許庁手法」の妥当性	253
8.7.1. 評価対象文数.....	253
8.7.2. 評価観点・評価基準の妥当性	253
8.7.3. 相対評価方法	254
8.7.4. 絶対評価方法	254
第9章 特許庁機械翻訳品質評価手法の改良と今後の課題	257
9.1. 評価精度の向上	257
9.1.1. 評価者間の意識合わせ方法の改善	257
9.1.2. 正解データの充実.....	259
9.1.3. 評価基準への補足.....	260
9.1.4. 流暢さの評価	264

9.1.5. チェックリストに基づく評価	265
9.2. 相対評価方法における課題	266
9.2.1. システムの定点観測	266
9.2.2. 相対順位の有意性が充分でない場合の対処策	267
9.2.3. 「流暢さの評価」におけるサンプル 50 文評価の妥当性	268
9.3. 絶対評価方法における課題	269
9.3.1. 「内容伝達レベルの評価」の精度の問題	270
9.3.2. 検証用データの精度の問題	270
9.3.3. 絶対評価スコアの定義に関する課題	271
9.3.4. 今後の方向性	272
9.4. テストセットに関する課題	273
9.4.1. パテントファミリー対訳文の利用	273
9.4.2. テストセットの再作成	273
9.4.3. 評価対象文が学習データに含まれている可能性の検証	274
9.5. 英日以外への言語対への適用可否	277
9.6. 自動評価	278
9.7. エンジン性能の総合的な評価	279
9.8. 検索システムを用いた評価	280

【添付資料】

添付資料 1-1. 特許庁における機械翻訳品質評価に関する基本的な考え方
添付資料 1-2. 仮設機械翻訳品質評価手法
添付資料 1-3. 特許庁機械翻訳品質評価手法
添付資料 1-4. 第 1 回委員会 議事内容及び主な提言・指摘事項
添付資料 1-5. 第 2 回委員会 議事内容及び主な提言・指摘事項
添付資料 3-1. 「英日機械翻訳の現状分析」サンプル文献一覧
添付資料 3-2. 「英日機械翻訳の現状分析」文献別重要技術用語一覧
添付資料 5-1. 英日テストセット データサンプル
添付資料 6-1. 品質評価作業用ファイル データサンプル
添付資料 7-1. 特許庁機械翻訳評価手法 評価作業手順書
添付資料 7-2. 「流暢さの評価」の評価サンプル
添付資料 7-3. 「内容の伝達レベルの評価」の評価サンプル
添付資料 7-4. 「重要技術用語の翻訳精度の評価」の評価サンプル
添付資料 7-5. 「チェックリストに基づく評価（言語共通項目）」詳細説明
添付資料 7-6. 「チェックリストに基づく評価（英日項目）」詳細説明
添付資料 7-7. 品質評価結果データ データサンプル
添付資料 9-1. 自動評価の粗ぶるいへの利用適否実験
添付資料 9-2. 機械翻訳エンジンにおける一般的なカスタマイズ機能例
添付資料 9-3. NTCIR 特許翻訳タスクにおける言語横断特許検索による Extrinsic 評価実験に関する調査

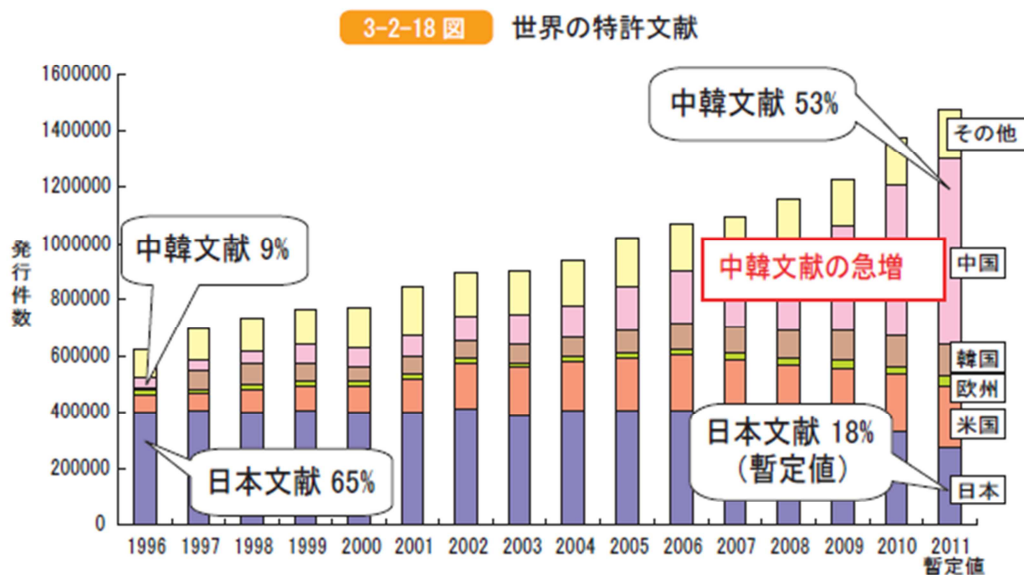
第 1 章

調査の目的と概要

第 1 章 調査の目的と概要

1.1. 調査の目的

特許庁では、従来より特許出願が多く、アクセスの必要性が高かった欧米の特許出願に加え、非欧米諸国の特許出願、特に中国からの出願が近年増大している状況を受け、中国語で記載された特許文献の要約について、人手翻訳により和文抄録を作成する事業を開始している。さらには、中国語で記載された特許文献の全文に対して機械翻訳を行い、翻訳結果を検索・照会可能とする中韓文献翻訳・検索システムのリリースを、平成 26 年度に予定している。



(出典) 特許行政年次報告書 2013 年版 特許庁

また、先般公表された「特許庁業務・システム最適化計画」改定版には、＜第Ⅰ期に実施する項目＞として、「中国・韓国語特許文献の翻訳検索環境を整備する。さらに、外国語特許文献のうち英語特許文献については、先行技術文献としての利用価値が高いことから、翻訳検索環境を整備する。」と記載されており、特許庁における外国語特許文献の機械翻訳及び人手翻訳の必要性は、ますます増大している。

こうした各種施策を行うにあたり、調達手続における技術審査や納入物の内容確認等において、機械翻訳結果の品質を適切に評価する必要があり、評価基準、評価用ツール、評価手順、及び評価体制等についての指針となる翻訳品質評価手法が求められる。

加えて、機械翻訳の分野においては、どの言語からどの言語への機械翻訳を行うかという組み合わせにより種々の特有課題及び留意点があると知られているが、英日機械翻訳に

における特有課題及び留意点、並びに現状における民間の英日機械翻訳の品質水準については、未知の部分が大きいのが現状である。

以上の状況を踏まえ、「特許文献機械翻訳の品質評価手法に関する調査」（以下、「本調査」という）では、特許庁における機械翻訳に関連する各種事業において機械翻訳結果の品質を適切に評価するための翻訳品質評価手法を調査・検証するとともに、現状における英日機械翻訳の問題点や品質等について調査し、特許庁における各種外国特許文献翻訳事業の翻訳品質の向上につなげることを目的とした調査を行う。

1.2. 調査内容

本調査では、翻訳品質評価手法の内容に関する調査を行うとともに、英日機械翻訳における品質水準、特有課題及び留意点の調査を行い、それらの調査結果、並びに有識者委員会（以下、「委員会」という）による議論を受け、「特許庁機械翻訳品質評価手法」を作成した。

また、英日機械翻訳に関するテストセットを作成するとともに、当該テストセットの一部について英日機械翻訳を実施し、当該英日機械翻訳結果について特許庁機械翻訳品質評価手法を実践し、その妥当性を検証した。

1.3. 調査方法の概要

本調査は、以下の方法により実施した。

※図 1-1（P.15）に、本調査の調査フロー図を示す。

1.3.1. 文献調査

「特許庁における機械翻訳品質評価に関する基本的な考え方」（添付資料 1-1）を踏まえ、書籍、論文、過去の特許庁調査における報告書及びインターネット情報等を利用して、以下の①及び②の内容に関する文献等を調査・整理・分析し、文献調査結果として取りまとめた。

① 翻訳品質評価手法の内容

翻訳品質評価に関する過去の主要な調査研究の内容、特許庁の実施した過去の調査における翻訳品質評価手法の内容を含む。

② 英日機械翻訳の内容

民間企業や研究機関等における主な英日機械翻訳の種類、品質水準、留意点等の内容を含む。

1.3.2. 英日機械翻訳の現状分析

文献調査の調査結果を踏まえ、英語特許文献 20 文献に対し 5 種類の英日機械翻訳エンジンによる機械翻訳を実施し、英日機械翻訳における技術用語の翻訳精度の評価を確認するとともに、英日機械翻訳における主な課題・留意点及び対応策を調査した。

1.3.3. 新設機械翻訳品質評価手法の作成

文献調査（1.3.1.）及び英日機械翻訳の現状分析（1.3.2.）の調査結果を踏まえて、仕様書に示された「仮設機械翻訳品質評価手法」（添付資料 1-2）の内容について、下記①～④の観点を含めて検討を行い、第 1 回委員会に諮るための案として「新設機械翻訳品質評価手法」を作成した。

- ① 評価手法が評価結果の利用目的（利用場面、利用方法）に沿っているか
- ② 評価手法が翻訳対象の特性に沿っているか
- ③ 評価手法が翻訳結果の利用方法に沿っているか
- ④ 評価手法が、客観的評価と主観的評価とでバランスがとれているか

1.3.4. 委員会による検討

本調査に関して専門的な視点からの検討、分析、助言を得るために、機械翻訳に関する専門家、翻訳品質評価に関する専門家、特許文献の翻訳に関する専門家等で構成される委員会を 2 度にわたり開催した¹。本委員会委員の一覧を以下に示す。

氏 名	所 属 ・ 専 門
委員長 江原 暉将	山梨英和大学 人間文化学部 教授（特任）
宇津呂 武仁	筑波大学 システム情報系 知能機能工学域 教授
後藤 功雄	日本放送協会（NHK） 放送技術研究所 専任研究員
山下 喜代治	一般財団法人工業所有権協力センター（IPCC） 調査業務センター 機械 B 部門 部門長
熊野 明	東芝ソリューション株式会社 プラットフォームソリューション事業部 ソフトウェア開発部 参事
浜口 宗武	株式会社知財コーポレーション 代表取締役
塚田 元	日本電信電話株式会社（NTT） コミュニケーション科学基礎研究所 主任研究員

¹ 第 1 回及び第 2 回委員会の主な提言・指摘事項は、添付資料 1-4「第 1 回委員会 議事内容及び主な提言・指摘事項」及び添付資料 1-5「第 2 回委員会 議事内容及び主な提言・指摘事項」を参照。

第1回委員会（2013年12月6日）では、「特許庁機械翻訳品質評価手法」の案として作成した「新設機械翻訳品質評価手法」の妥当性について検討した。

第2回委員会（2014年2月5日）では、第1回委員会での指摘を踏まえて作成した「特許庁機械翻訳品質評価手法（下記 1.3.5.参照）」について、同手法を用いて品質評価を実践（1.3.8.）した結果から、その妥当性、課題、問題点、ならびに他言語対への適用可否について検討した。

1.3.5. 特許庁機械翻訳品質評価手法の作成

第1回委員会の指摘事項等を踏まえて「新設機械翻訳品質評価手法」の内容を修正し、「特許庁機械翻訳品質評価手法」（添付資料 1-3）を作成した。

1.3.6. 英日機械翻訳に関するテストセットの作成

「特許庁機械翻訳品質評価手法」を実践するため、同手法にて定義したテストセットの要件に基づき、1,000 文からなる「英日機械翻訳に係るテストセット」を作成した。

同テストセットには、英日機械翻訳の現状分析（1.3.2.）にて調査した英日機械翻訳における特有課題及び留意点について検証可能な文を十分に含めた。

1.3.7. 英日機械翻訳の実施

作成した「英日機械翻訳に係るテストセット」から 200 文を機械翻訳対象文として抽出し、これらに対し 5 種類の英日機械翻訳エンジンによる機械翻訳を実施して、「特許庁機械翻訳品質評価手法」を実践するための「品質評価対象データ」を作成した。

1.3.8. 特許庁機械翻訳品質評価手法の実践

作成した「品質評価対象データ」に対し、「特許庁機械翻訳品質評価手法」による品質評価を実施して、「品質評価結果データ」（品質評価対象データに評価項目ごとの品質評価結果を付したもの）を取得した。

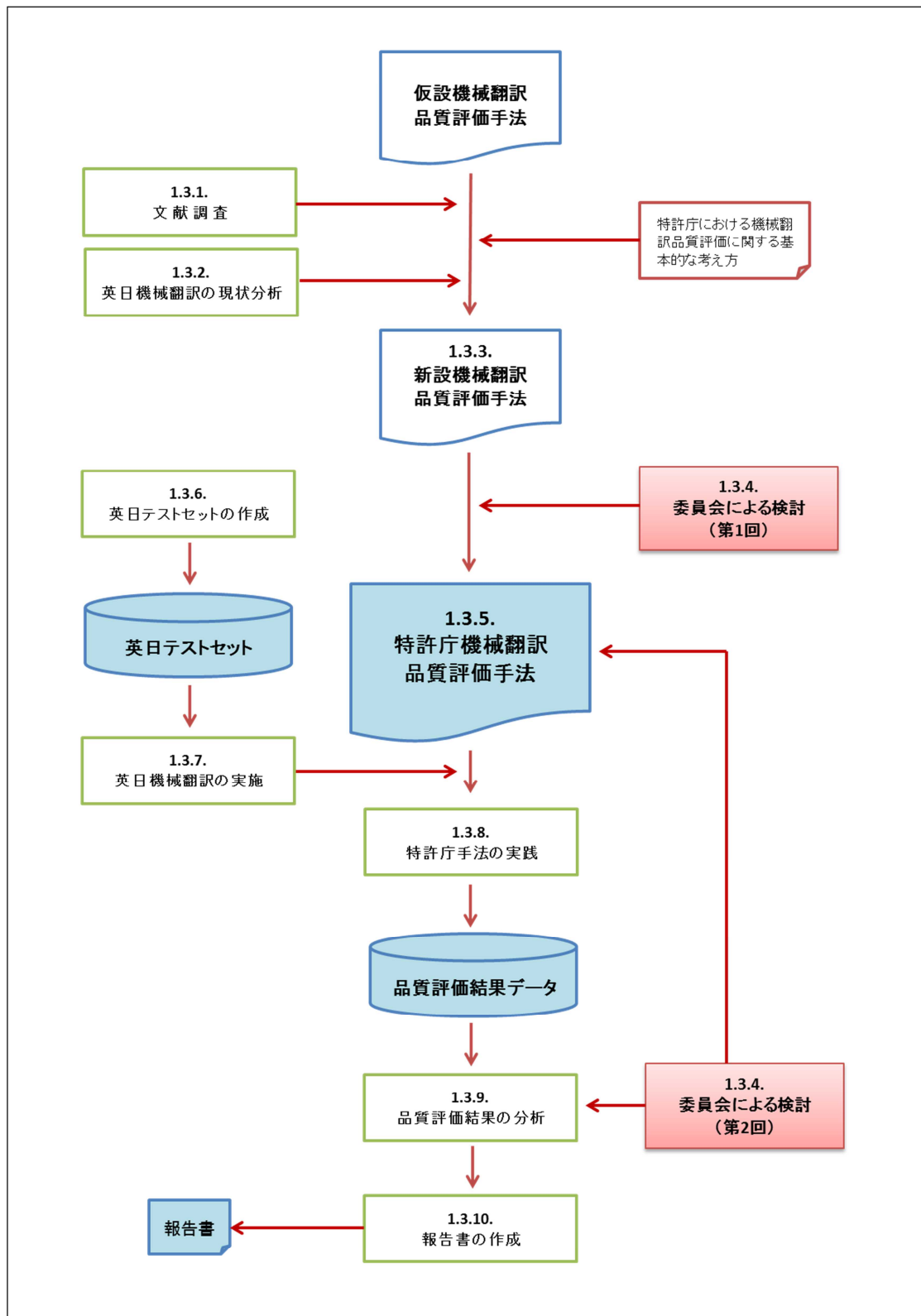
1.3.9. 品質評価結果の分析

1.3.8 にて取得した「品質評価結果データ」を用いて、「特許庁機械翻訳品質評価手法」の妥当性、課題、問題点、多言語への応用可否について分析した。

1.3.10. 報告書の作成

本調査の内容を報告書にまとめた（本報告書）。

図 1-1 「特許文献機械翻訳の品質評価手法に関する調査」調査方法フロー



1.4. 調査結果の概要

本項にて、本調査の調査結果の概要を示す。

1.4.1. 文献調査

本調査の実施にあたり、既存の主な翻訳品質評価手法の内容及び英日機械翻訳の内容について、各種文献等を利用して情報を収集した。以下に、文献調査の対象とした翻訳品質評価手法及び英日機械翻訳エンジン、そして調査の結果得られた知見・考察について示す。

1.4.1.1. 翻訳品質評価手法の内容

[機械翻訳品質評価に関する主要な先行研究]

- ・ ALPAC Report による評価
- ・ Mu プロジェクトによる評価
- ・ Adequacy と Fluency
- ・ WMT での評価
- ・ NTCIR での評価 (NTCIR-9/NTCIR-10 特許翻訳タスク)
- ・ 自動評価
- ・ テストセット
- ・ 誤訳内容の分類

[過去の特許庁調査における翻訳品質評価手法の内容]

- ・ 平成 20 年度「中国語機械翻訳技術に関する調査」
- ・ 平成 22 年度「特許庁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」
- ・ 平成 23 年度「中国特許文献の和文抄録作成に対する機械翻訳の活用に関する調査」

1.4.1.2. 英日機械翻訳の内容

[市販ソフトウェア (国内)]

- ・ The 翻訳プロフェッショナル V15 特許エディション (東芝ソリューション(株))
- ・ PAT-TranserV11 プロフェッショナル (株)クロスランゲージ
- ・ ATLAS V14 翻訳スーパーパック (富士通(株))
- ・ その他の市販ソフトウェア 5 製品

[市販ソフトウェア (海外)]

- ・ SYSTRAN 7 Premium Translator
- ・ TRADOS

[ウェブ上で利用可能なサービス]

- ・ 大手プロバイダによる無料サービス (Google、Yahoo!、Excite、@nifty)
- ・ EPO Patent Translate
- ・ 訳してねっと

[研究機関の英日機械翻訳エンジン]

- ・ NTT コミュニケーション科学基礎研究所 (SMT)
- ・ 山梨英和大学 (ハイブリッド)
- ・ 一般財団法人日本特許情報機構 (RBMT)

1.4.1.3. 文献調査から得た主な知見と考察

- ・ 機械翻訳結果の品質評価は、「内容の正確さ (Adequacy)」と「流暢さ (Frequency)」を評価観点として併用する方式が多く使われてきた。しかしながら、特許文献の機械翻訳結果においては「内容の正確さ」が最も重要であることから、「特許庁機械翻訳品質評価手法 (以下、「特許庁手法」という。)」は、「内容の正確さ」を観点とした評価をメイン評価とし、「流暢さ」は補助的な評価観点のひとつとして扱うのが適切である。
- ・ 「特許庁手法」は、「検索」用途、「粗読」用途をはじめ、特許文献の機械翻訳結果のさまざまな特許実務への有用性評価に用いるものである。したがって、特定の用途のみに特化した評価手法とするのは適当でなく、例えば各評価観点をモジュール化し、評価値を個別に算出、それぞれに重み付けする等、多様な用途に対応できるよう配慮する必要がある。
- ・ 「特許庁手法」のメイン評価となる「内容の正確さ」の評価は、主観的評価であるため、客観性の担保が求められる。客観的な評価が可能な手法として、機械翻訳結果に含まれる翻訳ミス抽出・分類し、その数を比較する方法が考えられるが、作業負担の大きさや難易度の高さから、この方法を必須とする評価手法は現実的でない。「特許庁手法」では、「内容の正確さ」の評価の客観性担保は、複数名での重複評価等、運用方法の工夫にて達成すべきである。ただし、「特許庁手法」の評価目的のひとつであるフィードバック用途に関しては、上記したような翻訳ミスの抽出・分類作業が必須となる。したがって「特許庁手法」には、フィードバック用途に限定した形でこの形式の評価手法も含める。
- ・ 過去の品質評価における評価対象文数は 150 件程度から数千件規模まで幅広く、どの程度の規模が妥当かの判断は難しい。文数が多いほど評価の信頼性は向上するが、費用対効果の観点からは、一定の信頼性を確保したうえで極力少量に抑えるべきである。「特許庁手法」では、評価対象文は 200 文に仮置きし、本調査にて実施する品質評価実践 (1.4.5.) にてその妥当性を検証する。
- ・ 現在、国内で市販されている主要な英日機械翻訳ソフトウェアはすべて RBMT 方式である。しかし、SMT 方式のエンジンも、Google 翻訳、EPO Patent Translate 等で実用化されており、NTCIR の特許翻訳タスクでも優秀な成績を収めている。「特許庁手法」は、こうした状況を踏まえ、多様な方式の機械翻訳結果を公平に評価できるような評価基準とする必要がある。

1.4.2. 英日機械翻訳の現状分析

「特許庁手法」の策定に先立ち、英日機械翻訳の現状分析のため、主要な英日機械翻訳エンジンにおける重要技術用語の翻訳精度の確認と、英日機械翻訳に特有の課題・留意点の洗い出しを行なった。

1.4.2.1. 英日機械翻訳における重要技術用語の翻訳精度の評価

英日機械翻訳エンジン 5 種類（RBMT×3、SMT×2）を用いて、日本語パテントファミリーを持たない英語特許文献 20 文献を機械翻訳し、各文献の機械翻訳結果に含まれる重要技術用語（1 文献あたり 10 語を選定）の翻訳品質を評価した。

〔評価基準〕

評価基準は、「特許庁手法」での採用を視野に入れ、以下の 4 段階を設定した。

- 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はないが、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。

〔評価方法〕

機械翻訳では、同一の用語であっても異なる訳語が出力される場合がある。このため、重要技術用語は出現の都度、評価した。その結果、重要技術用語 200 語（10 語×20 文献）に対し、計 3,161 回の評価が行われた。

〔評価結果〕

重要技術用語の翻訳精度評価の集計結果を表 1-1 に示す。表中の「語数」はのべの出現回数である。

表 1-1. 重要技術用語の翻訳精度評価 集計結果（サンプル 20 文献）

評価値		RBMT1	RBMT2	RBMT3	SMT4	SMT5
A 適訳	語数	2,011	2,117	2,194	2,394	2,264
	%	63.6%	67.0%	69.4%	75.7%	71.6%
B 可訳	語数	621	722	674	458	487
	%	19.6%	22.8%	21.3%	14.5%	15.4%
C 誤訳	語数	493	318	285	283	276
	%	15.6%	10.1%	9.0%	9.0%	8.7%
D 不訳	語数	36	4	8	26	134
	%	1.1%	0.1%	0.3%	0.8%	4.2%
総計	語数	3,161	3,161	3,161	3,161	3,161
	%	100%	100%	100%	100%	100%

「検索」用途への有用性は、重要技術用語が適訳語（A）に訳されているか否かが主要な評価観点となる。5種のエンジンにおける適訳語（A）の比率は、表 1-1 のとおり、SMT4 が最も高く 75.7%、RBMT1 が最も低く 63.6%であった。この結果から、現状の英日機械翻訳は、1 文献に含まれる技術用語のうち平均 6～7 割程度が適切に訳される水準であると考えられる。

一方、「粗読」用途や「精読」用途においては、可訳語（B）にも有用性があるため、適訳語（A）と可訳語（B）との合算値にて評価すべきである。下表 1-2 のとおり、適訳語と可訳語の合算値では、RBMT3 が最も高く 90.7%、RBMT1 が最も低く 83.2%であった。この結果から、現状の英日機械翻訳では、SMT と RBMT のどちらの方式のエンジンでも、標準的な難易度の文献であれば、文献中の 8～9 割の技術用語は正しい意味に訳されることが期待できる。

表 1-2. 重要技術用語の翻訳精度評価 A.適訳率+B.可訳率

評価値	RBMT1	RBMT2	RBMT3	SMT4	SMT5
適訳率	63.6%	67.0%	69.4%	75.7%	71.6%
可訳率	19.6%	22.8%	21.3%	14.5%	15.4%
合算値	83.2%	89.8%	90.7%	90.2%	87.0%

1.4.2.2. 英日機械翻訳の特有課題・留意点及び対応策

「特許庁手法」を策定するにあたり、英日機械翻訳における種々の課題や留意点をあらかじめ把握すべく、重要技術用語の翻訳精度の評価（1.4.2.1.）に用いた機械翻訳結果を分析し、英日機械翻訳における主な課題・留意点として、下記 16 項目を洗い出した。

<EJ-01. 文中のピリオドを適切に処理していない>

文中にてピリオドが記号の一部などで使われた際、これが誤って文区切りとして扱われる等、適切に処理されない場合がある。

<EJ-02. 「:」や「;」を用いた文を適切に処理していない>

英語特許文献では、クレーム部等にて、構成要素を一文中で「:」や「;」で区切って列挙する文が多用される。英日機械翻訳では、「:」や「;」を文末と解釈する場合があるが、その際には、前後の係り受けを適切に表現するための配慮が必要である。

<EJ-03. 重要な単数・複数情報が維持されていない>

英文では単数形と複数形の使い分けに大きな意味を持たせることができるが、そのような文において、機械翻訳結果がそのニュアンスを適切に表現できていない場合がある。

<EJ-04. 単数形・複数形併記表現を適切に処理していない>

英文では、要素が「単数および／または複数」である場合、これを単数形・複数形併記表現（例：an element and/or elements）やカッコ補記表現（例：element(s)）等で表現することがあり、これに適切に対応する必要がある。

<EJ-05. 序数詞表現が適切に翻訳されていない>

冠詞に関する特許英文頻出表現として、不定冠詞＋序数詞表現（a first XXX, a second XXX）や、無冠詞＋序数詞表現が多用されることが挙げられる。first、second 等の序数詞は通常定冠詞を伴うものであるため、エンジンによっては、こうした表現を想定していない場合がある。

<EJ-06. 慣用表現への語の挿入を適切に処理していない>

英文においては、複数の語から成る慣用表現の中間への語の挿入が一般的に行われるが、英日機械翻訳では、こうした語の挿入が原因で慣用表現をそれと認識できずに誤訳してしまう場合がある。

<EJ-07. 倒置文を適切に翻訳していない>

英語特許文献では、主語が長大になる場合に文意を理解しやすくする等の目的により、倒置文が多用される。

<EJ-08. 名詞と代名詞／省略表現の関係を適切に表現できていない>

同一文中にて既出の名詞が再度出現する場合、代名詞や省略表現（例：machine translation system であれば the system 等）が用いられることが多いが、文の構造によっては、機械翻訳結果にて両者の順序が入れ替わり、関係性がわかりにくくなる場合がある。

<EJ-09. 慣用表現とみなすべき場合／そうでない場合を適切に翻訳できていない>

同一の表現が慣用表現として使われている場合とそうでない場合とがあり、文意文脈に応じて正しく解釈する必要がある。

<EJ-10. 前置詞に対し適切な訳語を選択できていない>

of や for などの前置詞の適訳は、述語や目的語の属性等により変化する。例えば for であれば、述語や目的語に応じて「～のために」「～用の」「～という理由で」「～に向かって」など多数の訳語に訳し分ける必要がある。

<EJ-11. 接頭辞、接尾辞を伴う語を適切に処理していない>

英語特許文献では、任意の語の前後に接続して意味を追加する接頭辞・接尾辞が多用される。主な接頭辞・接尾辞の例としては、de-XXX、non-XXX、XXX-free、XXXly などが挙げられるが、こうした接頭辞・接尾辞の使用により、語が正しく翻訳されなくなる場合がある。

<EJ-12. スラッシュを用いた並列表現を適切に処理していない>

英語特許文献では、スラッシュ (/) を用いて名詞や動詞、前置詞等を並列させる表現も多用される (例 : is/are)。このようにスラッシュで連結された語が、適切に処理されない場合がある。

<EJ-13. therein、wherein、whereby 等を適切に処理していない>

英語特許文献では、「発明の名称」や「特許請求の範囲」をはじめ各所にて、therein, therefor, wherein といった自然文ではあまり用いられない副詞が多用される。こうした副詞を適切に表現する必要がある。

<EJ-14. 代名詞 same、the same を適切に処理していない (【発明の名称】)>

「発明の名称」部では、same や the same の代名詞用法が多用される。英日機械翻訳では、この用法に適切に対応できていない場合がある。

<EJ-15. 小文字を含む大文字表記を適切に処理していない (【発明の名称】)>

英語特許文献の「発明の名称」は全文字大文字で表記されることが多いが、まれに、全文字大文字表記でありながら、元素記号など一部のみ小文字が用いられるケースがある。このような特殊なケースが、機械翻訳処理の誤動作を招く場合がある。

<EJ-16. 引用クレームに関する表現を適切に翻訳していない (【特許請求の範囲】)>

「特許請求の範囲」において、前出クレームを引用するいくつかの定型表現 (例 : according to any preceding claim) がある。特許文献の英日機械翻訳では、こうした表現に適切に対応する必要がある。

これらの課題・留意点は、英日機械翻訳に関するチェックリスト項目として「特許庁手法」に反映させた。

1.4.3.「特許庁機械翻訳品質評価手法」の作成

本調査では、仕様書に示された「特許庁における機械翻訳品質評価に関する基本的な考え方」、並びに、先に実施した文献調査、英日機械翻訳の現状分析の調査結果を踏まえ、同じく仕様書に示された「仮設機械翻訳品質評価手法（以下、「仮設手法」という。）」の内容を再検討し、「特許庁機械翻訳品質評価手法（特許庁手法）」を作成した。

※「特許庁手法」の内容は添付資料 1-3 を参照のこと。

1.4.3.1.「特許庁手法」の前提条件

「特許庁手法」の前提条件は、以下のとおりとした。

[機械翻訳の対象]

機械翻訳の対象は、特許文献である。したがって「特許庁手法」の評価対象は特許文献の機械翻訳結果となる。ただし、文献単位ではなく、テストセットに含まれる単独の文単位での評価となる。

[機械翻訳結果の用途]

「特許庁手法」にて評価する機械翻訳結果の用途は、以下のとおりである。

- ① 検索用途：先行技術調査におけるテキスト検索用のデータとしての利用。
- ② 粗読用途：先行技術のあたりをつけるための粗読用データとしての利用。
- ③ 精読用途：先行技術の内容を正確に把握するための精読用データとしての利用。

機械翻訳結果は、特許庁内で使われている検索システムのコンテンツとして、上記の各用途に供され、ユーザ（審査官）は、基本的に日本語の文献と同様の手法によって本コンテンツを検索することを想定する。

[品質評価の目的・利用方法]

品質評価の目的・利用方法として、以下の 3 つを想定する。

- ① 相対評価：複数の機械翻訳システム間の翻訳精度比較や、特定のシステムの翻訳品質の定点観測（同一システムの二つの時点の翻訳精度比較）等。
- ② 絶対評価：評価対象としたシステムの特定の用途への有用性の判断等。具体的な用途として、検索、粗読のほか、実体審査での精読、権利侵害の存否等も想定される。
- ③ FB 利用：システムの具体的な弱点の把握と品質向上等のためのフィードバック利用。

「特許庁手法」は、このうち最も喫緊の目的である「相対評価」に主眼を置きつつも、「絶対評価」及び「フィードバック利用」にも十分に利用可能なものとする。また、「相対評価」

と「絶対評価」は1回の品質評価の結果が双方に利用されることも想定するが、「フィードバック利用」のための品質評価は、これらとは切り離して別途行われることを前提とする。したがって、「フィードバック利用」のための品質評価の手法は、「相対評価」「絶対評価」と共通である必要はない。

【評価者の前提】

「特許庁手法」の評価者は、特許文献とその利用に対する知識があり、かつ日本語に充分精通していることを前提とするが、高度な機械翻訳知識や文法知識、原文読解力は要件としない。

ただし、フィードバック利用のための評価に関しては、原文と機械翻訳結果とを照合して翻訳ミスの内容と原因を特定する作業が必須となるものと想定されるため、この評価用に設ける評価観点に限り、評価者の要件として原文読解力を前提とする。

また、「特許庁手法」は、評価者の人数・時間におけるリソースを考慮し、作業負担が過大とならぬよう配慮する。

1.4.3.2. 「特許庁手法」における主な検討事項

「特許庁手法（添付資料 1-3）」は、「仮設手法（添付資料 1-2）」をベースとし、各項目の定義内容を再検討する形で作成した。以下、主な検討内容を示す。

【事前準備】

「事前準備」部分では、評価用のテストセット（1,000 文程度）、評価対象文（テストセットから抽出した 200 文程度）、及び人手評価補助ツールについて定義した。

「仮設手法」は、対応特許から抽出した対訳文をテストセットに使用することを前提としていたが、対応特許文は機械翻訳エンジンのチューニング（SMT エンジンの学習データ等）に使用される可能性があり、その場合に正確・公平な品質評価が不可能となることから、「特許庁手法」では、対応特許の利用はあくまで選択肢のひとつとし、理想的には、対応特許不在の特許文に、人手翻訳で参照用翻訳文を付与してテストセットを作成すべきであることを明記した。

【評価の観点及び基準：(1) 内容の伝達レベルの評価】

「(1) 内容の伝達レベルの評価」は、「特許庁手法」に定めた 4 つの評価観点のひとつであり、機械翻訳結果が、原文の内容をどの程度正確に伝達しているかを 5 段階評価にて主観的に評価する。「特許庁手法」におけるメインの評価観点に位置づけられる。

「仮設手法」は、「内容の伝達レベルの評価」の際に「流暢さ」の要素を考慮するか否かが明確でなかった。「特許庁手法」では、「流暢さ」の評価は独立した観点として別途行う

こととし、「内容の伝達レベルの評価」では内容の正確さのみを評価するものとして、評価基準や定義内容を改めた。

また、「内容の伝達レベルの評価」は、主観的な評価であり、かつ複数人（3名以上）による重複評価であるため、評価者間の評価水準を合わせるため、①各評点へのパーセンテージの併記、②評価対象となる「重要情報」の定義、③評価者間の意識合わせの具体的な方法の定義、を追加した。

公平な相対評価を行うための要件として、複数の機械翻訳結果を評価する場合、同一の文については必ず全てのシステムの機械翻訳結果を同一の評価者が同時に評価することを要件とした（※他の評価観点についても同様）。

【評価の観点及び基準：(2) 重要技術用語の翻訳精度の評価】

「(2) 重要技術用語の翻訳精度の評価」は、「仮設手法」の「(2) 技術用語に関する誤訳の個数」の内容を見直したものである。テストセットの各文に含まれる重要技術用語の翻訳精度を評価するための観点であり、特に「検索」用途への有用性を測るために活用される。

「仮設手法」では、本観点の評価基準は「適訳（A）」と「誤訳（C）」の2段階であったが、「特許庁手法」では、機械翻訳結果に数多く含まれる「一般的に使われる訳語でなく同義の範囲外だが、正しい意味は分かる」という訳語を「可訳（B）」として「誤訳」と区別した。また、未知語、訳漏れについても、これを「誤訳」と区別し「不訳語（D）」とした。

【評価の観点及び基準：(3) 流暢さの評価】

「特許庁手法」にて追加した評価観点であり、機械翻訳結果の文としての読みやすさ、理解しやすさを評価する観点である。「流暢さ」は、特許文献の機械翻訳結果を評価するうえで主要な観点ではないものの、例えば「内容の伝達レベル」や「重要技術用語の翻訳精度」等が同程度の翻訳結果の優劣を定める際の補助的な指標のひとつとして必要である。

なお、「流暢さの評価」は、原文の内容にとらわれず、機械翻訳結果を独立した文として評価する必要がある。このため、評価の順序として、原文の内容を参酌する他の評価観点よりも先にこれを行うよう定めた。

【評価の観点及び基準：(4) チェックリストに基づく評価】

「特許庁手法」では、「チェックリストに基づく評価」は、機械翻訳エンジンの翻訳ミスの傾向や弱点を把握し、これをフィードバックしてエンジンの性能改善を促すために行う評価と定め、その用途に特化した内容とした。

「チェックリストに基づく評価」は、機械翻訳結果における言語共通の翻訳ミスをカテゴライズした【必須項目】（7項目）、同じく言語共通の課題ではあるが、特に機械翻訳が苦手とする特定のケースや、特許文献に特有または頻出のケースのみを対象とした【任意項

目】(12項目)、そして、特定の翻訳言語に特有の課題をまとめた【翻訳言語別任意項目】(英日：16項目)の3つのパートにて構成した。

【必須項目】には、機械翻訳結果に含まれる全ての翻訳ミス、必ず、かつ1項目のみにカウントすることとした。したがって、【必須項目】7項目のカウント値の合計が、その機械翻訳結果に含まれる翻訳ミスの総数となる。

【必須項目】にカウントされた各翻訳ミスは、その後【任意項目】及び【翻訳言語別任意項目】とも照合され、該当する全ての項目に重複カウントされる。なお、【任意項目】及び【翻訳言語別任意項目】は、項目数が多いことから、評価の目的等に応じて必要な項目のみを選定してよい。

また、「チェックリストに基づく評価」は、原文と翻訳結果を照合し、翻訳ミスの内容や原因を精緻に分析する必要があるため、原文の言語の読解力を備えることを評価者の要件とした。

【相対評価方法】

ここでは、「特許庁手法」を用いて実施した品質評価の結果を用いて、複数の機械翻訳結果を比較するための方法について定義した。具体的には、相対評価に使用する3つの評価観点について、それぞれの評価結果のスコア化の方法を定めた。

なお、「内容の伝達レベル」及び「流暢さ」は、評価対象文の文長により難易度が異なる性質のものである。このため、スコア化の際は、評価対象文(200文)における文長比率を、実際の特許文献の文長比率に補正することを定めた。

本調査では、英語文の文長を、語数を基準に「長文(51語以上)」「中文(26語～50語)」「短文(25語以下)」の3種に区分した。また、基準とする文長比率は、「英日機械翻訳の現状分析」に用いたサンプル20文献における比率に準拠し、長文1：中文4：短文5とした。

【絶対評価方法】

「絶対評価」とは、「特許庁手法」においては、機械翻訳エンジンの各種特許実務用途への有用性を品質評価結果から直接判断する評価を指す。特許実務用途は、「検索」用途、「粗読」用途のほか、実体審査での精読、権利侵害の存否等も含む。「相対評価」も基本的には特定の用途を想定して行うものであるが、ここで分かるのは品質評価スコアの比較による優劣のみであり、それぞれが各種用途にどの程度有用であるかは、「絶対評価方法」を用いて品質評価スコアと用途とをマッチングさせることで把握する。

「特許庁手法」では、各種特許実務用途への有用性(絶対評価スコア)を「内容の伝達レベルの評価」のスコアと「重要技術用語の翻訳精度の評価」の適訳率とを主な基準とし

で定めることとした。具体的には、「内容の伝達レベルの評価」において内容伝達スコアが 3.0 以上であれば「粗読」用途に有用、「重要技術用語の翻訳精度の評価」において適訳率が 60%以上であれば「検索」用途に有用とした。

【主要項目と補助項目】

「特許庁手法」には 4 種類の評価観点が定義されているが、評価作業のリソース（評価者の人数、時間等）の制限等により、全ての評価項目を要件どおりに実施することが困難な場合も考えられることから、これら 4 種類の評価観点を「主要項目（内容伝達、重要技術用語）」と「補助項目（流暢さ、チェックリスト）」とに区分し、全ての評価項目の実施が困難な場合は「主要項目」のみを実施可能とすることを許容した。

1.4.4. 英日機械翻訳に関するテストセットの作成

「特許庁手法」にて定めたテストセットの定義に基づき、英日機械翻訳に関するテストセット（以下「英日テストセット」という。）を作成した。ただし、同手法にて理想的とされる「対応特許の存在しない特許文献から選定し、人手翻訳にて参照用翻訳文を新規作成する」形での作成は作業期間的に困難であったため、選択肢のひとつとして許容される「パテントファミリー関係を利用して抽出した対訳文対」を用いてこれを作成した。

英日テストセットは 1,000 文にて構成し、「チェックリストに基づく評価」における【任意項目】や【翻訳言語別任意項目 [英日]】の各項目について検証可能な文を十分な量含めた。

1.4.4.1. 英日テストセットの構成

英日テストセットは、技術分野（IPC）、記載箇所、原文の長さ、文献種別等について、極力、実際の特許文献の状況に則した配分とした。表 1-3 及び 1-4 に、英日テストセットの構成を示す。

表 1-3. 英日テストセットの構成（IPC／記載箇所）

IPC セクション	A	B	C	D	E	F	G	H	合計
文 数	157 文	146 文	123 文	25 文	25 文	74 文	238 文	212 文	1,000 文
タイトル	8 文	7 文	6 文	1 文	1 文	4 文	12 文	11 文	50 文
要 約	8 文	7 文	6 文	1 文	1 文	4 文	12 文	11 文	50 文
クレーム	31 文	29 文	25 文	5 文	5 文	15 文	48 文	42 文	200 文
明細書	110 文	103 文	86 文	18 文	18 文	51 文	166 文	148 文	700 文
構成比	15.7%	14.6%	12.3%	2.5%	2.5%	7.4%	23.8%	21.2%	100%

表 1-4. 英日テストセットの構成（原文語数）

	短文（1～25 語）	中文（25～50 語）	長文（51 語～）	合計
文 数	500 文	400 文	100 文	1,000 文
構成比	50%	40%	10%	100%

1.4.4.2. 文の採用基準

1.4.4.1.にて定めた構成比率に従い、英日テストセットに採用する文を選定した。選定作業は文献単位で行い、候補データ中で最もテストセットに適すると思われる文を 1 文献あたり 1～2 文ずつ、以下の必須条件と優先条件を満たしているかを目視チェックしたうえで選定した。

- ① 原文の内容が文法的に正しい【必須】
- ② 原文と翻訳文の対応が正確である【必須、ただし翻訳文の修正は認める】
- ③ 特許文献の用途上有用である【必須】
- ④ チェックリスト任意項目の検証に適している【優先】

1.4.5. 特許庁機械翻訳品質評価手法の実践

「特許庁手法」の妥当性を検証するため、英日テストセットより 199 文を抽出し、これを「英日機械翻訳の現状分析」に用いたのと同じ 5 種類の機械翻訳エンジン（RBMT×3、SMT×2）で機械翻訳した結果に対して「特許庁手法」の各観点による品質評価を実践した。

※今回の品質評価実践は、「特許庁手法」の妥当性の検証を目的としたものであり、評価対象とした各エンジン間の翻訳条件は公平ではない。したがって、本調査にて作成した「品質評価結果データ」をそのまま各エンジンの性能比較に利用することは推奨しない。

1.4.5.1. 「内容の伝達レベルの評価」

「内容の伝達レベルの評価」は、「特許庁手法」の定義に従い、3 名の評価者による重複評価にて実施した。3 者の評価結果は、以下に示す手順で「評価スコア（表 1-5）」とした。

- ① 各文について 3 評価者の評価値の平均値を算出
- ② 全文について、長文、中文、短文ごとの平均値²を算出
- ③ 上記各平均値を実際の特許文献の文長比率³に補正して「評価スコア」を算出

表 1-5. 「内容の伝達レベル」評価スコア

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
スコア	2.39	2.79	3.01	3.15	2.76
相対順位	5	3	2	1	4

上記評価スコアに基づく RBMT1～SMT5 の相対順位は、「正解」としてあらかじめ作成していたリファレンス評価結果と完全に合致し、また、3 名の評価者別の評価スコアも、1 名のみ RBMT2 と SMT4 の順位に逆転があったものの、それ以外は全てこれと同一であった。このことから、「特許庁手法」の「内容の伝達レベルの評価」は、評価者による相対順位のばらつきがほとんど生じないことが確認された。

また、評価スコアの符号検定の結果、エンジン間の相対順位の大半において、十分な有意性が確認された。このことから、「特許庁手法」における「200 文程度」という評価対象文数の設定は、（極めて性能の接近したエンジンが対象で、かつ本評価観点において必ず優劣を定めなければいけない、という特殊な条件でない限り）妥当な規模であることがわかった。

² 本調査では、英語文の文長を、語数を基準に「長文（51 語以上）」「中文（26 語～50 語）」「短文（25 語以下）」の 3 種に区分した。品質評価実践の対象 199 文における文長比率は、長文：10 文、中文：88 文、短文：101 文とした。

³ 「英日機械翻訳の現状分析」に用いたサンプル 20 文における比率に準拠し、長文 1：中文 4：短文 5 とした。

したがって、「特許庁手法」の定める「内容の伝達レベルの評価」が、複数のエンジン間の相対評価に十分に実用できるものであることが検証された。

ただし、主に絶対評価において問題となる評価水準（評価の厳しさ、甘さ）に関しては、3名の評価者間で大きなずれが生じていることが判明した（各評価者の平均スコアは2.75、3.32、2.47）。その原因として、今回の品質評価実践では、作業期間や作業体制の都合により、適切な形で評価者間の事前意識合わせを実施できなかったことが挙げられる。また、評価者が基準とする評価サンプルが不足していたことも指摘されており、対応が必要である。

1.4.5.2. 「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の評価」は、199文それぞれに対してあらかじめ選定された最大3語の重要技術用語、計524語を対象として実施した。下表1-6に評価結果を示す。

表 1-6. 「重要技術用語の翻訳精度」評価結果

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
評価 A (適訳語)	273 語 52.1%	317 語 60.5%	348 語 66.4%	443 語 84.5%	394 語 75.2%
評価 B (可訳語)	108 語 20.6%	87 語 16.6%	90 語 17.2%	41 語 7.8%	46 語 8.8%
評価 C (誤訳語)	136 語 26.0%	110 語 21.0%	83 語 15.8%	39 語 7.4%	61 語 11.6%
評価 D (不訳語)	7 語 1.3%	10 語 1.9%	3 語 0.6%	1 語 0.2%	23 語 4.4%
相対順位 (適訳率)	5	4	3	1	2

上記結果に関しても符号検定を行い、全ての相対順位において十分な有意性が検定された。この結果から、重要技術用語の翻訳精度の評価という観点において、200文、1文あたり最大3語という「特許庁手法」の設定は、相対評価を行う上で妥当なデータ数であることが検証された。

1.4.5.3. 「流暢さの評価」

「流暢さの評価」の評価スコアは表 1-7 のとおりとなった。こちらの相対順位も、符号検定の結果、大半において十分な有意性を有することが確認された。

表 1-7. 「流暢さ」評価スコア

	MT1	MT2	MT3	MT4	MT5
スコア	2.46	2.78	3.02	3.16	2.83
相対順位	5	4	2	1	3

1.4.5.4. 「チェックリストに基づく評価」

「チェックリストに基づく評価」についても、199 文を対象に、【必須項目】【任意項目】【翻訳言語別任意項目 [英日]】それぞれの全チェック項目について実施した。表 1-8 に、【必須項目】の集計結果を示す。

表 1-8. 「チェックリスト評価」【必須項目】カウント集計結果

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
A 語の欠落	18	13	18	32	100
B 不要語の混入	11	4	11	18	26
C 未知語	9	11	11	5	30
D 品詞誤り	83	59	46	55	55
E 係り受け誤り	164	127	129	140	158
F 語の誤訳	420	355	270	119	162
G その他	30	14	20	42	29
合 計	735	583	505	411	560
(少ない順)	5	4	2	1	3

「チェックリストに基づく評価」は、評価対象となるシステムの翻訳ミスの傾向や弱点を把握する目的で行われるものであるが、上記カウント集計結果から、例えば SMT5 では「A. 語の欠落」が他のエンジンに比べて非常に多いことや、RBMT1 では品詞誤りが他のエンジンに比べて多めであること等、いくつかの傾向を読み取ることができる。ただし、これらは全て複数のエンジンの集計結果の比較によって判明したことであり、「チェックリストに基づく評価」は、仮に評価対象とするシステムが 1 種のみであったとしても、比較用として複数のシステムを同時に評価することが推奨される。

なお、【任意項目】【翻訳言語別任意項目 [英日]】の評価結果も同様の傾向を示した（集計結果は省略）。

1.4.6. 「特許庁手法」の実用性と今後の課題

「特許庁手法」を用いた品質評価実践の結果や、有識者委員会での指摘に基づき、同手法の実用性と、今後の課題及び対応策について考察した。

1.4.6.1. 「特許庁手法」の実用性

〔複数エンジンの相対評価への実用性〕

本調査にて作成した「特許庁手法」は、同手法を用いて実施した品質評価実践（1.4.5.）により、複数エンジンの相対評価に充分に実用できることが検証された。評価対象文数についても、エンジンの性能が極めて接近したものでない限り、「特許庁手法」に定義された200文で有意性の高い相対順位が得られることが分かった。

〔特定システムの定点観測〕

「特許庁手法」は、相対評価の一種である「特定のシステムにおける翻訳品質の定点観測（継時的変化の確認：同一システムの二つの時点の翻訳精度比較）」に用いられることも想定される。

特定システムの定点観測も、新旧の機械翻訳結果をその都度作成して相対評価する場合は、本調査で実施した複数システム間の翻訳精度比較と大きな違いはない。一方、新システムのみを新たに評価して、旧システムについては既存の評価結果を使用するという方式を採る場合は、新旧システムで同一の評価対象文を用い、評価者に旧システムの機械翻訳結果とその評点を提示することが有効であろう。評価者は、旧システムの機械翻訳結果と、それに対して付与された評点を参照し、これとの比較で新システムの評価を行うことができるため、正確な相対評価が可能となり、また評価水準もおのずと前回水準と揃うこととなる。

〔英日以外への言語対への適用可否〕

本調査は、英日機械翻訳を対象として「特許庁手法」の作成及び品質評価実践を行なった。ただし、本手法自体は、言語対を限定せず汎用的に適用できるものとなるよう配慮して作成している。これを検証するため、本調査では、中日機械翻訳結果（RBMT 及び SMT、計4種類）を対象に、「特許庁手法」による品質評価を試行し、問題なく適用可能であることを確認した。

ただし、「チェックリストに基づく評価」の【翻訳言語別任意項目】は、本調査では英日機械翻訳に関するもののみを設定しており、他の言語対の品質評価に「チェックリストに基づく評価」を含める場合は、それに先立ち【翻訳言語別任意項目】を設定する必要がある。

1.4.6.2. 「特許庁手法」の課題

〔評価水準のずれの解消〕

「内容の伝達レベルの評価」において、3名の評価者間で評価水準に大きなずれが発生した。その主な原因は、評価期間や評価体制の都合上、評価者間の十分な意識合わせができなかったためと考えられる。これを踏まえ、「特許庁手法」では、事前意識合わせの方法について具体的に示し、これを要件化した。

また、品質評価実践結果から、評価者に提示するサンプルや、評価者間の意識合わせ等に用いるための「正解データ」163文を作成した。ここから精選したものを評価サンプルに補充することで、今回の品質評価実践において評価サンプルが不十分であったという可能性への対処とする。

〔絶対評価方法の妥当性検証〕

「特許庁手法」に定義した絶対評価方法の評価基準は、今回の品質評価実践にて得た評価結果と、別途用意した検証用データとを照合して、両者に整合するよう調整したものである。

ただし、今回の検証は、単一の検証データ（1名の元特許庁審査官による絶対スコア判定結果）のみに基づいたものであり、したがって上記評価基準の妥当性は十分に検証されたとはいえない。今後、より多数かつ精度の高いデータを取得して、さらにその妥当性を検証し、基準を調整していく必要がある。

〔テストセット〕

本調査にて作成した英日テストセットは、英日の対応特許文献から対訳文を抽出して作成した。しかしながら、こうした対応特許文献の対訳文は一般に入手可能な既存データであるため、例えば SMT 方式の機械翻訳エンジンにおける学習データ等の形でエンジンのチューニングに利用される可能性がある。この場合、テストセットに含まれる対応特許文をチューニングに利用したエンジンが不当に有利な条件で品質評価されることとなり、評価結果の信頼性が低下する。本調査における品質評価実践でも、評価対象とした一部エンジンにて、評価対象文がチューニングに利用された形跡が見られた。

SMT 方式の機械翻訳エンジンは既に各所で実用化されており、「特許庁手法」を用いた品質評価においても、今後常に評価対象に含まれることが予測される。これを考えると、テストセットは全文、目的言語で記載された対応文献（英日機械翻訳であれば日本語文献）が存在しない特許文献から抽出し、参照文を人手翻訳にて作成することが推奨される。

1.4.6.3. 特許文献機械翻訳結果の評価に関する課題

ここでは、特許文献の機械翻訳の評価全般に関する課題について考察する。

[自動評価の利用について]

「特許庁手法」を用いた人手による品質評価に先立ち、自動評価手法を利用して、人手評価を行うまでもない低品質の機械翻訳結果をあらかじめ検出・除外（粗ぶるい）することが可能であるか実験した。その結果、今回実験に用いた BLEU、NIST、RIBES とも、人為的に作成した低品質の機械翻訳結果に対し、他よりも顕著に低いスコアが与えられた。

ただし、これら 3 種の自動評価には、SMT 方式を RBMT 方式よりも高く評価する傾向が見られた。このため、自動評価を利用した粗ぶるいの際には、方式間の有利不利を考慮する必要がある。各自動評価とも、同一の方式内での相対順位にはある程度の信頼性があると思われるため、同一の方式内で他よりも顕著に低いスコアが与えられたエンジンを粗ぶるいするという方法が考えられる。

[エンジン性能の総合的な評価]

本調査の品質評価実践は、あくまで「特許庁手法」の妥当性を検証するためのものであり、評価対象としたエンジンの大半はほぼ初期設定で利用している。しかしながら、今後実際に機械翻訳エンジンを相対評価する場合は、各エンジンのカスタマイズ機能等を利用して、それぞれを特許文献に最適化し、実用環境に近づけた設定で比較評価するべきである。

また、「特許庁手法」は機械翻訳結果そのものの品質を評価するための手法であるが、機械翻訳エンジンの性能は、現時点での機械翻訳結果の品質以外にも、こうしたカスタマイズ機能の幅広さや利用の容易さ、さらには各種の支援機能等、将来的な翻訳品質の改善可能性という観点からも考慮されるべきである。さらに、エンジンの優劣評価には、インタフェースの利便性や価格、ユーザ側でコントロール可能な要素の範囲など、翻訳品質以外にも多数の観点が存在し、機械翻訳エンジンの相対評価にあたっては、翻訳品質だけでなく、これらの要素も観点に加える必要がある。

第 2 章

文 献 調 査

第 2 章 文献調査

本調査を進めるにあたり、その予備知識として、「翻訳品質評価手法の内容」及び「英日機械翻訳の内容」について、各種文献を利用して調査した。本章にその調査結果をまとめる。

2.1. 翻訳品質評価手法の内容

「特許庁機械翻訳品質評価手法（特許庁手法）」の作成に先立ち、既存の主要な翻訳品質評価手法の内容について、書籍、論文、特許庁における過去の調査報告書、およびインターネット情報等を利用して調査した。以下にその調査結果を示す。

2.1.1. 機械翻訳品質評価に関する主要な先行研究

本節では、これまでの機械翻訳の歴史において機械翻訳結果の人手による評価がどのように行われてきたか、機械翻訳評価の主要な先行研究について記す。

2.1.1.1. ALPAC Report による評価 [1][3]¹

ALPAC Report は、米国の National Academy of Sciences National Research Council が ALPAC (Automatic Language Processing Advisory Committee) を組織して行なった最初の大規模な機械翻訳利用全般の評価に関する報告書であり、1966 年に発行されたものである。報告書の正式名称は “Language and Machines—Computers in Translation and Linguistics —” であるが、ALPAC Report として知られている。この報告書では、科学技術文献を対象にしたロシア語から英語への機械翻訳システムを評価している。評価の結果、機械翻訳の利用は有効性がないと結論付けられ、その後の機械翻訳の研究、特に米国での研究に多大な影響を与えた報告書である。

ALPAC Report で評価された機械翻訳結果は 144 文であり、Intelligibility (理解容易性) と Fidelity (忠実性) という 2 観点から評価されている。Intelligibility は訳文が理解しやすいかどうかを評価する基準である。そのため、原文のロシア語の表現とは切り離れた訳文の英語のみが評価される。一方、Fidelity は訳文が原文の情報を忠実に反映しているかどうかを評価する基準である。ただし実際には、Fidelity を直接評価せず Informativeness (情報付加性) として評価している。この評価法は、まず、訳文の英語を読んでそこに含まれる情報内容を知り、その後、原文のロシア語を読んで、訳文では得られなかった情報がどの程度付加されるかで評価するものである。したがって評価者は英語とロシア語の双方が理解可能な者となる。Intelligibility は表 2-1 に示す 9 段階評価 (9 が最良) で評価され、Informativeness は表 2-2 に示す 10 段階評価 (機械翻訳の良さとしては 1 が最良) で評価される。

¹ 本章にて見出しの末尾等を示す角カッコ表記は、章末に示す参考文献に対応する。

表 2-1. Intelligibility のスケール[1]

スケール	説 明
9	Perfectly clear and intelligible. Reads like ordinary text; has no stylistic infelicities. 完璧に明瞭で理解可能である。自然な文として読むことができ、文体に不適切な箇所がない。
8	Perfectly or almost clear and intelligible, but contains minor grammatical or stylistic infelicities, and/or mildly unusual word usage that could, nevertheless, be easily “corrected.” 文全体又はその大部分が明瞭で理解可能であるが、文法・文体の軽微な誤り、及び／又は、中程度の不適切な用語用法を含む。ただし、これらは容易に「訂正」が可能である。
7	Generally clear and intelligible, but style and word choice and/or syntactical arrangement are somewhat poorer than in category 8. 全般的に明瞭で理解可能であるが、文体、用語、及び／又は構文において、スケール 8 より若干劣る。
6	The general idea is almost immediately intelligible, but full comprehension is distinctly interfered with by poor style, poor word choice, alternative expressions, untranslated words, and incorrect grammatical arrangements. Postediting could leave this in nearly acceptable form. 文の大意はすぐに理解可能だが、不適切な文体や用語、一般的でない表現、未知語、文法的に不正確な構文により、完全な理解は明らかに妨げられる。
5	The general idea is intelligible only after considerable study, but after this study one is fairly confident that he understands. Poor word choice, grotesque syntactic arrangement, untranslated words, and similar phenomena are present, but constitute mainly “noise” through which the main idea is still perceptible. 文の大意を理解するには相当量の思考が必要だが、その結果、文の大意を把握したことは確信できる。不適切な用語、奇妙な構文、未知語等が存在するが、その大半は「ノイズ」のレベルであり、文の主旨は依然として把握可能である。
4	Masquerades as an intelligible sentence, but actually it is more unintelligible than intelligible. Nevertheless, the idea can still be vaguely apprehended. Word choice, syntactic arrangement, and/or alternative expressions are generally bizarre, and there may be critical words untranslated. 一見、理解可能な文に見えるが、実際は理解不可能に近い。ただし、文の主旨は漠然と把握できる。用語、文体、表現等が全体的に奇妙であり、重要な語が未知語となっている場合もある。
3	Generally unintelligible; it tends to read like nonsense but, with a considerable amount of reflection and study, one can at least hypothesize the idea intended by the sentence. 全体的に理解困難である。意味不明な文となりがちであるが、相当量の熟考を用いれば、少なくとも文意の推測は可能。
2	Almost hopelessly unintelligible even after reflection and study. Nevertheless, it does not seem completely nonsensical. 相当量の熟考を用いても、理解はほぼ絶望的である。ただし、完全に意味不明とまではいえない。
1	Hopelessly unintelligible. It appears that no amount of study and reflection would reveal the thought of the sentence. 絶望的に理解不可能であり、どのような熟考を経ても文意の把握は不可能と思われる。

表 2-2. Informativeness のスケール[1]

(This pertains to how informative the original version is perceived to be after the translation has been seen and studied. If the translation already conveys a great deal of information, it may be that the original can be said to be low in informativeness relative to the translation being evaluated. But if the translation conveys only a certain amount of information, it may be that the original conveys a great deal more, in which case the original is high in informativeness relative to the translation being evaluated.／評価対象の翻訳結果を読解後に原文を参照し、どの程度新たな情報が得られるか、すなわち原文の「情報付加性／Informativeness」を評価する。翻訳結果が原文の内容を十分に伝達できていれば、その後に原文を参照しても新たに得られる情報は少ないため、翻訳結果に対する原文の情報付加性は低くなる。逆に、翻訳結果が原文の内容を僅かしか伝達していなければ、原文から新たに得られる情報は多くなる。この場合、翻訳結果に対する原文の情報付加性は高くなる。)

スケール	説 明
9	Extremely informative. Makes “all the difference in the world” in comprehending the meaning intended. (A rating of 9 should always be assigned when the original completely changes or reverses the meaning conveyed by the translation.) 原文の情報付加性が極めて高い。文意を理解するうえで「雲泥の差」がある。(スケール9は、翻訳結果から得られた情報が原文により完全に変化又は反転するような場合に付与する。)
8	Very informative. Contributes a great deal to the clarification of the meaning intended. By correcting sentence structure, words, and phrases, it makes a great change in the reader's impression of the meaning intended, although not so much as to change or reverse the meaning completely. 原文の情報付加性が高い。文意の明確な把握に大きく寄与する。文構造、用語、フレーズが訂正されることで、把握していた文意が大幅に変更される。ただし、文意を覆すほどの変更ではない。
7	Between 6 and 8 スケール6と8の間
6	Clearly informative. Adds considerable information about the sentence structure and individual words, putting the reader “on the right track” as to the meaning intended. 原文が明らかに情報付加性を有する。文構造や用語に関して相当量の情報が追加され、文意が正しく把握できるようになる。
5	Between 4 and 6 スケール4と6の間
4	In contrast to 3, adds a certain amount of information about the sentence structure and syntactical relationships; it may also correct minor misapprehensions about the general meaning of the sentence or the meaning of individual words. スケール3と異なり、文構造や統語関係に関する若干の情報が追加される。これにより、文の大意や個々の用語に対する軽微な誤解が正される場合がある。

3	By correcting one or two possibly critical meanings, chiefly on the word level, it gives a slightly different “twist” to the meaning conveyed by the translation. It adds no new information about sentence structure, however. 主に語レベルで、1～2箇所を正しい意味に訂正できる。これにより、翻訳結果の伝えていた意味が若干変化する。ただし、文構造に関しては新たな情報の追加はない。
2	No really new meaning is added by the original, either at the word level or the grammatical level, but the reader is somewhat more confident that he apprehends the meaning intended. 実質的な新情報は語レベルでも文法レベルでも得られないが、把握していた文意が正しいことについての確信が得られる。
1	Not informative at all; no new meaning is added, nor is the reader’s confidence in his understanding increased or enhanced. 全く情報付加性がない。得られる新情報はなく、把握していた文意が正しいことについての確信にも変化がない。
0	The original contains, if anything, less information than the translation. The translator has added certain meanings, apparently to make the passage more understandable. 原文よりも、むしろ翻訳結果のほうが情報量が多い。翻訳者が何らかの情報を付加しており、これによってより理解し易い文になっている。

2.1.1.2. Mu プロジェクトによる評価 [2][3]

Mu プロジェクトは 1980 年代に日本で行われた日英と英日の機械翻訳システム研究開発のプロジェクトであり、科学技術文献の速報への利用を目指したものである。このうち、参考文献[2]で評価しているのは日英システムであり、評価対象文数は 1,682 文である。ここでの評価基準は、理解容易性と忠実度、および修正容易度と誤解積度である。このうち理解容易性は ALPAC Report の Intelligibility、忠実度は Fidelity にほぼ相当する。修正容易度と誤解積度は、機械翻訳結果に対して後修正を施し正確な訳文とする場面を想定した評価であり、修正容易度は、外国人（英語を母国語とする平均的英米人）による訳文の理解に誤りがない場合において後修正のしやすさを評価し、誤解積度は外国人による訳文の理解に誤りがある場合において後修正の負担を評価するものである。これらの評価項目は次のように定義される。

- 1) 理解容易性：出力の訳文が、どの程度理解できるか、の観点からの評価。日英翻訳では、英語を母国語とする平均的英米人（以下では『外国人』と呼ぶ）が、原文（日本語文）を見ずに出力の訳文だけを見てどれだけ理解できるか、を評価する。
- 2) 忠実度：出力文が入力文の持つ情報をどれだけ忠実に反映しているか、出力文と入力文との間で文意のズレがどれほどか、に関する評価。
- 3) 修正容易度：表 2-5 に示す 4 段階（後修正が必要な中でランク A が最良）で定義される。
- 4) 誤解積度：表 2-6 に示すステップで作業したときの 4 段階評価（誤解釈があった中でランク P が最良）で定義される。

理解容易性は表 2-3 に示す 5 段階評価（1 が最良）である。評価に揺れが生ずることを防止するために具体的な判断基準を示している。例えば、評価者が修正すべきと考える箇所に校正用の記号を記入するが、評価値 2 と 3 との境界は「外国人（評価者）が独力で自信を持って校正用記号を記入できるかどうか」で判断される。また、3 と 4 の境界は「出力文に校正用記号を入れる気になるかどうか」で判断される。

表 2-3. 理解容易性の基準[2]

基 準	説 明
1	文意は明瞭で疑う余地はない。文法、用語用法、文体共に適切で、何ら修正の必要はない。
2	文意は明瞭で理解できるが、文法、用語用法、文体上に多少の問題が見られ、1 には至らない。ただし、こうした欠陥を改善するにあたっては出力文だけを見て外国人評価者単独で修正を施すことができ、日本人評価者の援護を必要としない。
3	全体的文意は把握できるが、文法、用語用法上の問題のため細部理解に自信が持てない。こうした問題を取り除き、さらに文体面の改善を施そうとすると、修正方法がひとつに定まらず、入力文の意味するところを日本人評価者に確認しなければならない。
4	文法、用語用法上の問題が多いので、相当の思考の後わずかに文意を仮定できるか、もしくはほとんど把握できない。こうした問題を解決して文体を改良するよう手を加えるより、新たに翻訳をやり直した方が時間も速く、質も向上すると思われる。
5	全く理解できない。どんなに熟考、検討しても文意をとることは不可能である。

忠実度は表 2-4 に示す 7 段階評価（0 が最良）であり、次のステップに従って評価される。

ステップ 1：入力文は日本語として自然な文か。そうでなければ、入力文の不備として評価の対象としない。

ステップ 2：出力文は文として成立しているか。文構成上重要な要素（主語、動詞）が欠如し、文として成立していない時は、評価点を 6 とする。

ステップ 3：出力文に、節や句（名詞句）の欠落がある場合は、評価点を 5 にする。

ステップ 4：語の欠落、節と節、節と句（関係節の修飾先）、名詞句と述部、の係り受け関係に誤りがある場合、評価点を 4 とする。

ステップ 5：名詞句内の修飾関係、時制・数・態・副詞の位置、に誤りがある場合、また、不必要に同じ単語が出力されている場合、評価点を 3 とする。

ステップ 6：理解容易性が 3～5 の場合、評価値を 2、理解容易性が 2 の場合、評価値を 1、理解容易性が 1 の場合、評価値を 0 とする。

表 2-4. 忠実度の基準[2]

基 準	説 明
0	入力文の構造は出力文に忠実に再現されていて、外国人が見ても明瞭で、修正する必要はない。
1	入力文の構造は出力文に忠実に再現されていて、また外国人にも明瞭に理解できるが、多少の修正は必要である。ただし、その修正は与えられた出力文だけにに基づき、外国人が単独で行える。
2	入力文の構造は出力文に忠実に再現されているが、語の置き換えを必要とする。
3	入力文の構造は出力文に忠実に再現されているが、句と句の関係、時制、数、態、副詞の位置等に誤りがある。また、単語が重複して出力される等の誤りがある。
4	入力文の構造は出力文にあまり良く再現されていない。句の要素の欠落、節と節、節と句、あるいは節内の関係に誤りがある。
5	入力文の構造は出力文に良く再現されていない。節や句の欠落があるが、一応文としては成立している。
6	入力文の構造は全く出力文に再現されていない。主語や述部の欠落があるため、文として成立していない。ただし、表題のような名詞句にあっては、その主名詞（日本語で最後に位置する名詞）の欠落またはこれを直接修飾する節・句の動詞的要素が欠落している。

修正容易度は表 2-5 に示す 4 段階（ランク A が最良）で評価する。

表 2-5 修正容易度の基準[2]

ランク	説 明
A	外国人が自らその修正案を提示して、日本人翻訳者が『yes』と答えるだけで、修正方法が決定できる場合
B	最初の修正案が『yes』とはならず、それに代わる修正案を外国人が 3 回を限度に提示する。その 3 回までの中に日本人翻訳者が『yes』と答えられる案があった場合
C	以上の過程では適切な修正が発見できない場合、今度は日本人翻訳者の側から 3 回を限度に修正案を出し、それで外国人が文意を諒解できた場合
D	以上の過程で適切な修正案が発見できなかった場合

誤解程度は表 2-6 に示す作業手順で決定する（誤解釈があった中でランク P が最良）。

表 2-6. 誤解程度の評価手順[2]

ステップ	説 明
1	外国人が修正した個所または無修正で通過した個所について、入力文との意味の差が大きいと日本人翻訳者が判断した場合、その旨を外国人に指摘し、外国人も指摘通り検討の必要を認めた場合、誤解釈があったと判断する。
2	日本人翻訳者が修正指摘を行い、外国人が『yes』と答えれば、ランク P の誤解釈とする。
3	指摘に対して『yes』が成立しなかった場合、それに代わる代案を 3 回を限度に提示し、それで解決すればランク Q、それでも解決しなかった場合はランク R の誤解釈とする。

参考文献[2]では、上記の各評価項目と評価基準に基づいて実際に評価を行い、最後に評価について今後改良が必要な点として以下のことを指摘している。

- 1) 理解容易性を「平均的英米人」が訳文をどれだけ理解できるか、で定義したが、「平均的」の定義があいまいである。今後、種々の背景知識を持った外国人の評価結果を比較検討することにより、個人の影響がどの程度あるかを明らかにする必要がある。
- 2) 今回の評価では、すべて 1 文を対象とした。したがって、文脈の影響は全くない状態での評価になっている。今後、抄録²単位での評価が不可欠である。
- 3) 原文の難しさを何らかの形式で定量化し、これとの関係で理解容易性や忠実度を論じなければならない。
- 4) 修正容易性の評価は、機械翻訳の利用形態と関係して変化する。理解容易性の尺度も変化するであろう。

² ドキュメント単位のこと。Mu プロジェクトでは科学技術文献速報での抄録の翻訳を対象にしているので抄録単位となる。

2.1.1.3. Adequacy と Fluency[4]

ALPAC Report から 35 年後の 2001 年からスタートした米国国防総省の DARPA (The Defense Advanced Research Projects Agency) の TIDES (Translingual Information Detection, Extraction and Summarization) プログラムで、中英とアラビア語英語の機械翻訳結果の評価が行われた。ここで用いられたのが Adequacy と Fluency である。この評価基準は、その後、NIST (National Institute of Standards and Technology) の OpenMT (Open Machine Translation) Evaluation プログラムで用いられ、今日でも広く用いられている。これらの評価基準は、以下の質問に対する答えとして定義される。答えは、それぞれ表 2-7 及び表 2-8 に示す 5 段階評価であり、両者とも 5 が最良である。

Adequacy : How much of the meaning expressed in the reference translation is also expressed in the hypothesis translation?

(リファレンス翻訳文に示される意味内容が、機械翻訳結果にどの程度示されているか?)

Fluency : How do you judge the fluency of this translation?

(機械翻訳結果の流暢さはどの程度か?)

これらの基準は翻訳先言語として英語を仮定しており、評価者はまず Fluency を評価し、その後 Adequacy を評価する。

表 2-7. Adequacy の評価基準[4]

基 準	説 明
5	All (全て)
4	Most (ほとんど)
3	Much (かなり)
2	Little (若干)
1	None (無し)

表 2-8. Fluency の評価基準[4]

基 準	説 明
5	Flawless English (完璧な英語)
4	Good English (良好な英語)
3	Non-native English (非ネイティブ英語)
2	Disfluent English (流暢でない英語)
1	Incomprehensible (理解不可能)

なお、DARPA の TIDES プログラムにおける評価対象データ数は、中英が 878 セグメント³（文相当）、アラビア語－英語が 728 セグメントである。

2.1.1.4. WMT での評価[5][6]

WMT（Workshop on Statistical Machine Translation）は 2006 年から毎年行われているヨーロッパ言語を対象とした機械翻訳のワークショップである。2013 年には、英語とスペイン語、フランス語、ドイツ語、チェコ語、ロシア語との間の機械翻訳が対象となっている。評価対象文は各言語対について 3,000 文である。

WMT における機械翻訳結果の評価は人手評価と自動評価である。初期の WMT では人手評価として Adequacy と Fluency を用いていた。その際、評価値の不整合を除くために、評価者間と評価文間で、評価値のばらつきに対する正規化⁴を行っている[5]。その後、表 2-9 に示すようにさまざまな人手評価が試みられたが、現在は文レベルの Ranking が用いられている。

表 2-9. WMT における人手評価の変遷

年	Adequacy	Fluency	Ranking (文レベル)	Ranking (セグメントレベル)	Yes/No 判定 (セグメントレベル)	後編集 + Yes/No 判定
WMT06	○	○				
WMT07	○	○	○	○		
WMT08			○	○	○	
WMT09			○			○
WMT10			○			○
WMT11			○			
WMT12			○			
WMT13			○			

WMT では、一つのテスト文に対して 5 システム分の出力を評価する方式をとっている。Ranking はこの 5 システムの出力を 1 位から 5 位までの間でランクづけするものであり（同一順位を複数のシステムにつけることは認められている）、Adequacy や Fluency のような絶対評価ではなく相対評価である。テスト文とシステムとの組み合わせを変えることで全システムに対する評価値を決めているが、相対評価であることには変わりがない。Adequacy と Fluency の評価をやめた理由としては、これらの評価基準が曖昧で信頼性が低いことが挙げられている。実際、評価者間での評価値の違いを分析した結果、異なる評価

³ 実際のデータが入手できず推測となるが、データが文だけでなく名詞句やフレーズを含んでいるため「セグメント」という表現が用いられたものと考えられる。

⁴ 評価者間の評価のばらつきを除くために、各評価者の生の評価結果の平均値を 3 から減じた値を生の評価値に加え正規化している。各評価文間の評価のばらつきを除くために、各評価文の生の評価結果の平均値を 0 から減じた値を生の評価値に加え正規化している。

者が同一翻訳結果に対して同一評価値を与える確率は Fluency で 40%、Adequacy では 38% であった[6]。

2.1.1.5. NTCIR での評価

NTCIR (NII Testbeds and Community for Information access Research) は、国立情報学研究所が主体となって進めている情報検索、質問応答、要約、テキストマイニング、機械翻訳などに対する研究フォーラムを形成するプロジェクトであり、1.5 年を 1 サイクルとしてワークショップが行われている⁵。特許文献の機械翻訳に関しては、NTCIR-7 にて「特許機械翻訳タスク」として対象に加えられ、NTCIR-10 まで実施された。2007 年から 2008 年にかけて行われた NTCIR-7 では、人手評価の基準として Adequacy と Fluency が用いられた[7]。NTCIR-8 では人手評価は実施されなかったが、NTCIR-9 と NTCIR-10 では Adequacy と Acceptability が用いられた。また NTCIR-10 では、審査官による審査実務への利用を想定した評価も行われている。NTCIR-9/NTCIR-10 の特許機械翻訳タスクにおける人手評価については、次節 2.1.2.にて詳述する。

2.1.1.6. 自動評価

人手評価の実施には相応のリソース（時間、コスト等）が必要であることから、これを代替する自動評価が提案された。初期に提案されたものとしては、逆翻訳を用いるもの[8]や BLEU[9]が挙げられる。後者は、人間が作成した理想的な翻訳結果（Reference）とシステムの出力とを比較して、Reference に近い出力を良い出力とみなすものであり、今日でも広く用いられている。その後、NIST[10]、IMPACT[11]、RIBES[12]など多くの自動評価基準が提案され、現在もより良い自動評価基準を目指して研究が続けられている。しかしながら、自動評価と人手評価との相関は、統計方式機械翻訳（SMT）の翻訳結果のみを評価した場合にはかなり高いものの、SMT とルールベース方式（RBMT）が混在する場合は低くなる。RBMT の翻訳結果に対して、人手評価は高い評価値となり、自動評価は低い評価値となる傾向がある[13]。参考文献[7]では、複数の Reference を用意することで、この傾向をある程度回避できると述べているが、総じて自動評価は未だ発展途上であるといえる。

2.1.1.7. テストセット

機械翻訳システムの基本的な翻訳性能を評価するために、テスト文と、そのテスト文を機械翻訳した結果に対する評価の観点を記述したテストセットが提案・公開されている[14, 15]。これらのテストセットでは、機械翻訳における誤訳を語彙の誤りと構造の誤りとに大別し、それぞれについて評価項目を細かく分類して対応するテスト文を作成している。また、現在は公開されていないが、英日翻訳用のテストセットも JEIDA（社団法人日本電子

⁵<http://ntcir.nii.ac.jp/jp/about/>より引用

工業振興協会) により作成されている。JEIDA 英日翻訳用テストセットのテスト文と評価観点の一例を示す。

【例文】 There is an egg on the table.
【訳文】 テーブルに卵がある。
【質問】 “an egg” が「卵」のように、不定冠詞 “an” の訳語が省略されて訳されていますか？

しかしながら、これら既存のテストセットは、機械翻訳システムの基本的な翻訳性能の評価のために作成されたものであり、上例のとおりテスト文はシンプルなものが大多数である。このため、語彙・構文ともに複雑、かつ特有の表現を多く含む特許文に対する翻訳性能を評価するには最適とはいえない。特許文献の機械翻訳結果を、その使用目的に沿って評価するためには、専用のテストセットを新たに用意する必要がある。

2.1.1.8. 誤訳内容の分類

機械翻訳の出力を評価する際、**Adequacy** や **Fluency** のように出力文全体を総合的に評価するのではなく、そこに含まれる個々の誤訳内容に踏み込んだ分類評価が行われることがある。例えば、参考文献[16]においては、英語とスペイン語間および中国語から英語の機械翻訳を対象に誤訳内容の分析を行っている。

機械翻訳の誤訳は、語彙の誤りと構造の誤りに分けられる⁶。参考文献[16]では、語彙の誤りを(a)Missing Words、(b)Incorrect Words、(c)Unknown Words、(d)Extra Words の4種類、構造の誤りを(e)Word Order、(f)Punctuation の2種類にそれぞれ分類し、さらにこれらを細分類して詳細に誤訳内容を分類している。しかしながら、現実の文に対する機械翻訳結果は、これらの誤訳が複合的に生じ、特に、英語と日本語や中国語と日本語のように構文構造が大きく異なる言語間では、誤訳内容を分類することが難しい場合がある。例えば、意味不明の訳語が出力された場合、この訳語に対応する原文の語がどの語であるかの判定も難しく、さらに語順が大幅に入れ替わっているため、単なる **Incorrect Words** なのか **Word Order** の誤りも含むのかは判断が難しい。原語知識を有し、かつ機械翻訳に関する専門知識を有する者であればこのような判断も可能であるが、最終的な訳文だけしか見ることができない一般利用者にとっては、誤訳の詳細な分類は困難な場合も多い。

⁶より抽象的には、以下のように記述することができる[17]。自然言語処理における計算機内部での言語表現は、ラベル付き有向グラフとして捉えられる。そこで機械翻訳での誤訳は、このグラフの構造の誤りとラベルの誤りに分けられる。ここでは、前者を構造の誤り、後者を語彙の誤りとした。この分類に従えば、品詞付けの誤りは語彙の誤りに含まれる。

2.1.2. NTCIR-9/NTCIR-10 での人手評価

本節では、2.1.1.5.で言及した NTCIR-9 特許機械翻訳タスク[18]および NTCIR-10 特許機械翻訳タスク[19]での人手評価について詳しく説明する。

NTCIR-9/NTCIR-10 では、機械翻訳の用途は「情報収集のための利用」に設定されており、機械翻訳結果を次の条件において評価している。

- (1) 後編集を前提としない。
- (2) 翻訳結果から原文の内容を理解したい。

これらの用途のための評価基準として、Adequacy、Acceptability、特許審査評価という 3 つの評価基準が用いられている。NTCIR-9/NTCIR-10 特許機械翻訳タスクにおいて実施された人手評価を表 2-10 に示す。

表 2-10. NTCIR-9/NTCIR-10 特許機械翻訳タスクにおいて実施した人手評価

	Adequacy	Acceptability	特許審査評価
NTCIR-9	✓	✓	
NTCIR-10	✓	✓	✓

以下、評価で利用された Adequacy と Acceptability の評価基準及び評価実施方法、評価基準ごとの課題、テストデータの構築方法、評価に必要な文数の検討、及び特許審査評価について説明する。

2.1.2.1. Adequacy

Adequacy の主な目的は、システム間の比較である。NTCIR-9/NTCIR-10 においては、Adequacy の評価は文単位で、翻訳結果における翻訳原文中の意味の保持度合いを 5 段階で評価している。表 2-11 に Adequacy の評価基準を示す。表中、5 段階のグレード設定 (None ~ All) は参考文献[20]に基づく。この評価基準は、評価レンジが広く、低い訳質から高い訳質まで分類できるという特徴がある。また、客観的な基準ではないため、訳質の低いシステムばかり、あるいは高いシステムばかりを評価する場合は、それらの訳質の差が明らかになるように 5 分類するという運用が可能である。NTCIR-9/NTCIR-10 では、実質的に相対評価として運用されている。訳語選択が適切かだけでなくフレーズや節の意味が正しいかも考慮される。

表 2-11. Adequacy 評価基準

基 準	説 明
5	All (全て)
4	Most (ほとんど)
3	Much (かなり)
2	Little (若干)
1	None (無し)

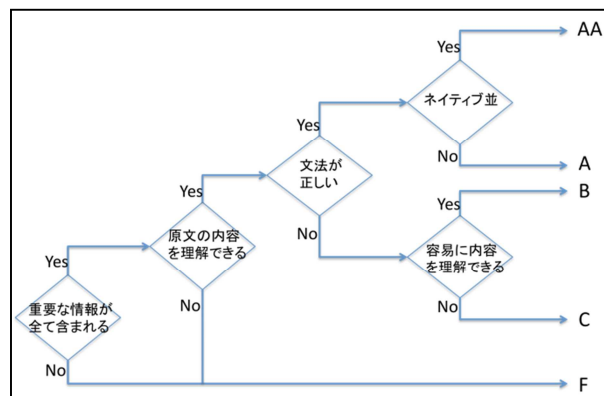
2.1.2.2. Acceptability

Acceptability の主な目的は、原文の意味が理解できる文数の割合を明らかにすることである。Acceptability の設計方針は次のとおりである。

- ・各評価グレードに意味を持たせる。例えば、「原文の意味が理解できる」や「原文の意味が理解できて文法が正しい」などを評価グレードとして用いる。
- ・流暢さも同時に評価する。原文の意味が理解できる場合に限り、流暢さも評価する。これによって、原文の内容を反映しているかどうかを無視した評価を避けられる。

Acceptability の評価は文単位で、文レベルの意味を評価する。図 2-1 に Acceptability の評価基準を示す。Acceptability は、入力文の意味が正しく伝わらない（例えば、重要な情報が一つでも欠けている）場合は全て F 評価となる。この評価は、Adequacy と比べて、より実用に近い評価を目指したものである。例えば、システムへの要求水準が「入力文の意味が分かればよい」であれば、C 以上の評価を得た訳文が有用であり、システムへの要求水準が「入力文の意味が分かり、かつ文法的に正しい」であれば、A 以上の評価を得た訳文が有用となる。このように、要求水準に応じた訳質の文の割合を明らかにすることができる。Adequacy 評価の結果からは、このような文の割合は把握できない。

図 2-1. Acceptability 評価基準



2.1.2.3. 文の訳質の評価基準の課題に関する考察

ここで、NTCIR-9/NTCIR-10 で文の訳質の評価基準として使用された Adequacy と Acceptability、そしてこれらに加え、一般的な評価基準のひとつであり Adequacy とペアで評価に用いられることが多い Fluency について、それぞれの評価基準としての課題を考察する。

[Adequacy の課題]

基準に客観性がなく、評価の各グレードが実際にどのような訳質であるか明確でない。例えば、原文の意味が理解できる文の割合は不明である。また、湧き出し語⁷に対する定義がない。目的言語⁸側での表現の流暢さに対する評価である Fluency は別評価とされているが、Adequacy と Fluency の役割分担が明確でない。

[Fluency の課題]

Fluency は流暢さのみを評価するため、原文の意味を反映していない流暢な訳文に高いスコアが与えられる。しかし、原文の意味を反映している訳文のみに流暢さのスコアを与えなければ、訳質の評価にはならない。例えば、入力文の日本語が「今日は晴れです。」で、出力の英語が「Hello!」の場合、誤訳であり訳質は明らかに低いですが、目的言語側での流暢さだけを考えれば全く問題点がないため、Fluency は最高評価となる。このような誤訳に高いスコアを付与しても翻訳の評価として適切でないことは明らかである。情報収集のために特許文献を機械翻訳した結果を評価する場合、原文の意味が反映されていない翻訳結果に高いスコアを付与することは、目的から考えて不適切である。そのため NTCIR-9/NTCIR-10 では Fluency の評価は実施せず、代わりに訳文の流暢さも評価基準に含む Acceptability 評価を実施している。

[Acceptability の課題]

中／低品質な訳文に対する評価の解像度が低い。例えば、一部誤訳でも全部誤訳でも最低評価になってしまう。このため、現状の機械翻訳の精度を考えると、システム間の比較は Adequacy の方が向いていると考えられる。

また、実際の情報収集の用途での有用性とのギャップがある。部分的にしか正しく訳せていない文でも、重要な部分が分かれば有用である場合がある。しかし、Acceptability 評価では、部分的にしか正しく訳せていない文は最低評価になってしまう。そのため、部分的に正しく訳せている翻訳結果についても、有用性を評価できることが望ましい。ただし、なぜこの文を翻訳する必要があるのかという目的を明確にしなければ、部分的に正しく訳せている文の有用性を評価することは困難である。

⁷ 翻訳結果のみに存在し、原文にその対応語が存在しない語を指す。

⁸ 翻訳結果の言語、ターゲット言語。英日翻訳であれば日本語が目的言語となる。

2.1.2.4. Adequacy/Acceptability 評価の実施方法

NTCIR-9/NTCIR-10 では、各機械翻訳システムあたり 300 文が評価された。評価者（有償）は 3 名で、各評価者は 1 システムあたり 100 文を評価した。すなわち 3 人の評価者は、それぞれ異なる入力文の評価を担当している。システム間で公平な評価となるよう、入力文ごとに全ての評価対象の翻訳結果を比較できる状態で同時に評価している。[18][19]

なお、評価コストを抑えるため、評価者には特許の各分野の専門家ではない者が起用された。このため、ここでは専門用語の訳語の厳密な適切さは評価困難となる。しかしながら、現状の機械翻訳の目標レベルは「原文にどのようなことが書いてあるのかが理解できるレベル」であると考えられ、評価はこの目標レベルの訳質に対するものである点、そして多くのシステムにおいて、原文の意味を保持できていない出力文が多い点等を考慮すると、専門用語の訳語をさほど厳密に評価せずともシステム間の比較評価は可能であり、評価基準に沿った評価はほぼ実施できているといえる。

また、NTCIR-9 の中英翻訳の評価以外は、目的言語が母国語の評価者が評価を実施している。NTCIR-9 中英翻訳の評価は、母国語が中国語で英語も理解できる評価者が評価を実施している。評価者の事前訓練として、3 人の評価者が同一の訓練用の 50 文（または 100 文）について評価を実施し、基準を合わせるために評価者が話し合っ統一の評価結果をこの 50 文（または 100 文）に対して決め、この統一の評価結果を評価の目安として用いている。[18][19]

入力文の意味の把握は、NTCIR-9 では、入力言語と目的言語のどちらも理解できる評価者が、入力文を読んで把握している。NTCIR-10 では、参照訳を読んで把握している。参照訳を使って評価する場合は、使わない場合に比べて次の利点と欠点があると考えられる。

- ・利点：目的言語のみを理解できれば評価できるため、評価を担当できる人材を見つけやすく、単価も入力文を理解する必要がある場合と比べて多くの場合に安くできる。専門用語の正解訳の例が得られるため、その正解訳例と一致した場合には、専門家でなくても専門用語の訳が適切であることが分かる。
- ・欠点：正解訳に引きずられて評価にバイアスがかかる可能性がある。訳語は参照訳とは異なるが、意味はだいたい分かるような訳でも、訳語が一致していないだけで低く評価する評価者が出てくる可能性がある。NTCIR-10 では 30 文ほど評価が進んだところで評価結果を集計して検査し、意味はおおむね分かるが参照訳と訳語が一致していないために低い評価になっている例を抽出して、必要に応じて評価の見直し作業を実施している。

2.1.2.5. テストデータの構築方法 [18] [19]

訓練データ⁹は 2005 年以前の特許文献から構築されているため、機械翻訳用のテストデータはこれらを含まない 2006～2007 年の特許文献から構築されている。NTCIR-9 におけるテストデータの構築方法は、自動的に抽出した対訳文対からランダムに文対を選択し、さらにその中から人手で正しい対訳の文対を選択するというものである。一方、NTCIR-10 のテストデータの構築方法は、対訳のある原言語の特許文献からランダムに文を選択し、それらの参照訳は新たに翻訳して構築するというものである。参照訳には、対訳の特許文献中の専門用語が使用されている。NTCIR-9 の構築方法は、対訳文対を特許文献から自動的に抽出する際に実際の特許文の傾向が完全には反映されなくなる可能性がある。それに対して NTCIR-10 の構築方法は、NTCIR-9 の構築方法に比べてより実際の特許文の傾向が反映されると考えられる。

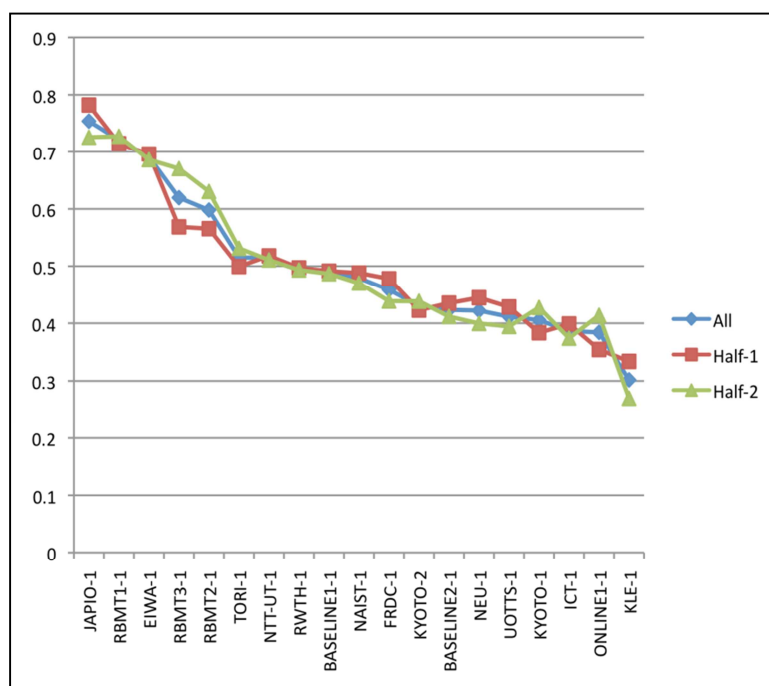
2.1.2.6. 評価に必要な文数に関する考察

文単位の評価でシステムを比較する場合において、信頼できる評価を得るために必要な文数についての検討が必要となる。もしテストデータが非常に少量（例えば 1 文のみ）であれば、たまたまその文だけうまく翻訳できた、またはできなかったという可能性が高く、システムの比較評価の信頼性は高いとはいえない。したがって、信頼性のある評価にするためには、ある程度の文数の訳質を評価する必要がある。

NTCIR-9 では、前述のとおり、各システムあたりランダムに選択した 300 文を評価しているが、この文数の妥当性を検証するため、評価結果をシステムごとに前半 150 文と後半 150 文に分けて比較を行い、双方における Adequacy によるシステム間の比較がどの程度一致するかが調査された[18]。システム間の比較には、正規化のために、Adequacy の値ではなくシステム間の相対評価値（詳細は参考文献[18]の pairwise score を参照）に換算した値が用いられた。日英翻訳の調査結果を図 2-2 に示す。図中、横軸はシステム ID を示している。Half-1 は前半の 150 文の評価結果、Half-2 は後半 150 文の評価結果、All は 300 文の評価結果を示している。

⁹ 機械翻訳システムのチューニングのために利用されることを想定して参加者に配布されるデータ。

図 2-2. 異なる 150 文でのシステム比較（日英翻訳）



また、表 2-12 に、日英、英日、中英での異なる 150 文でのシステム比較の Pearson 相関係数を示す。この相関係数は、Half-1 と Half-2 のシステムの相対評価値の相関を示している。

表 2-12. 異なる 150 文でのシステム比較の Pearson 相関係数

	Pearson 相関係数
日英	0.940
英日	0.985
中英	0.963

図 2-2 より、Half-1 と Half-2 の間では、多少のスコアの違いはあるが上位のシステムと下位のシステムが入れ替わるような大きな違いはないこと、また、表 2-12 より、日英だけでなく、英日、中英の翻訳結果においても、異なる 150 文のテストデータによる評価結果において、システム間の比較で高い相関を示していることが分かる。このことから、システム間の評価の差が大きい場合、150 文でのシステム間の比較結果の信頼性は高いといえる。したがって 300 文の場合はそれ以上の信頼性があると考えられる。

2.1.2.7. NTCIR-10 での特許審査評価 [19]

NTCIR-9 におけるランダムに選択した文の訳質評価により、現在の最高性能の機械翻訳システムは、特許文（クレーム部分を除く）の翻訳において半数以上の文で原文の内容を理解できることが確認された。この結果から機械翻訳が実用上においても有用である可能性が高いと考えられ、特許翻訳が必要とされる状況において機械翻訳を利用した場合にどれだけ有用であるかという観点における評価の 1 つとして、NTCIR-10 特許機械翻訳タスクにて「特許審査での有用性に基づく評価（特許審査評価）」が実施された。以下、その内容を記す。

特許審査官は、審査において既存の特許を調査して同じ技術が存在すれば、その特許文献を引用して審査対象の出願を拒絶する。そのため、既存の特許文献に記載された技術的内容である事実を認定する必要がある。既存の特許が外国語で書かれていてその言語が分からない場合には、翻訳する必要がある。この翻訳を機械翻訳で行った場合を想定し、機械翻訳結果から引用文献の特許を認定するために有用な事実をどれだけ認定できるか、という観点に基づいて、機械翻訳の有用性が評価された。

[全体の評価の流れ]

全体の評価の流れは下記(1)～(3)のとおりである。

- (1) 準備：評価で利用するデータを構築する。まず審査の結論が拒絶の審決（審査の最終決定）とその審査で引用された特許を取得する。そして、審査において「引用文献から審査官が認定した事実」の根拠となる文を引用文献から抽出する。抽出した文をテストデータとする。なぜなら、この抽出した文は、「審査官が認定した事実」を表しているためである。そして、このテストデータが正しく翻訳されれば、翻訳結果から「審査官が認定した事実」を認定することができる。
- (2) 翻訳：テストデータを機械翻訳する。
- (3) 評価：翻訳結果から、「引用文献から審査官が認定した事実」をどれだけ認定でき、審査に有用であるかについて評価する。

[評価基準]

特許審査評価で用いられた評価基準を表 2-13 に示す。有用性の評価は、過去の審査で審査官が引用文献の特許から認定した事実を、引用文献を機械翻訳した結果からどれだけ認定できるかに基づいて行われた。なお、評価は引用文献単位で行われた。

表 2-13. 特許審査評価の評価基準

VI	引用発明を認定するために有用な事実が全て認定できて、翻訳結果のみで審査可能
V	引用発明を認定するために有用な事実が半分以上認定できて、審査に有用
IV	引用発明を認定するために有用な事実が 1 つ以上認定できて、審査に有用
III	IVに至らないが、部分的に事実が認定できて、その文献が審査で無視できないことが分かる
II	一部の事実が認定できたが、審査に有用とはいえない
I	全く事実が認定できず、審査の役に立たない

(引用文献単位の評価)

[実施方法]

特許審査評価は日英翻訳と中英翻訳に対して行われた。特許庁審査官の経験があり、英語の能力が高い 2 名の評価者が起用された。1 システムあたり各評価者が 20 文献、合計 40 文献（文献の重複を含む）を評価した。評価したシステムは、**Adequacy** の結果が上位のシステムから手法の多様性を考慮して日英、中英それぞれ 3 システムが選択された。

評価者からは、前記評価基準に基づく評価を実施した後、各システムに対する総合評価としてのコメントも採取された。特許審査評価においては、実際に特許審査実務を経験した者が機械翻訳結果を役に立つと思うかどうかを最も重要なポイントであり、評価者からのコメントはこのポイントを直接的に表現することができるため、システムの有用性の判断に有用であると思われる。

[データの構築]

特許審査評価に必要なデータは、審決を用いて以下のように構築された。

- (1) 結論が不成立（拒絶）の審決を取得する。
- (2) 審決中に記載されている「引用文献から審査官が認定した事実の説明」を抽出する。
- (3) 審査官が認定した事実を構成要素単位に分けて、それぞれの内容の根拠となる文を引用文献から抽出する。抽出した文を機械翻訳で翻訳するテストデータとする。

表 2-14 に、審査官が認定した事実と引用文献から抽出した文の例を示す。表中左列は、審決中に記載されている審査官が引用文献から認定した事実であり、中央列は、審査官が認定した事実を構成要素単位に分けたものである。右列は、中央列の事実を認定する根拠となった引用文献中の文を抽出したものであり、この文が機械翻訳されるテストデータである。29 文献のテストデータを構築している。テストデータの構築は、弁理士が行なっている。中英翻訳のテストデータは、日英翻訳のテストデータを日中翻訳して構築している。なお、文献[19]で指摘されているように、テストデータは翻訳して構築するのではなく、実在のデータを用いることが望ましいであろう。

表 2-14. 審査官が認定した事実と引用文献から抽出した文の例

審査官が認定した事実	構成要素単位に分けた 審査官が認定した事実	引用文献から抽出した文 (テストデータ)
これらの記載事項によると、引用例には、 「内部において、先端側に良熱伝導金属部 43 が入り込んでいる中心電極 4 と、 中心電極 4 の先端部に溶接されている貴金属チップ 45 と、 中心電極 4 を電極先端部 41 が碍子先端部 31 から突出するように挿嵌保持する絶縁碍子 3 と、 絶縁碍子 3 を挿嵌保持する取付金具 2 と、 中心電極 4 の電極先端部 41 との間に火花放電ギャップ G を形成する接地電極 11 とを備えたスパークプラグにおいて、 中心電極 4 の直径は、1.2～2.2mm としたスパークプラグ。」 の発明が記載されていると認められる。	内部において、先端側に良熱伝導金属部 43 が入り込んでいる中心電極 4	また、図 3 に示すごとく、中心電極 4 の内部においては、上記露出開始部 431 よりも先端側にも良熱伝導金属部 43 が入り込んでいる。
	中心電極 4 の先端部に溶接されている貴金属チップ 45	また、中心電極 4 の先端部には、貴金属チップ 45 が溶接されている。
	中心電極 4 を電極先端部 41 が碍子先端部 31 から突出するように挿嵌保持する絶縁碍子 3	上記中心電極 4 は、電極先端部 41 が碍子先端部 31 から突出するように絶縁碍子 3 に挿嵌保持されている。
	絶縁碍子 3 を挿嵌保持する取付金具 2	上記絶縁碍子 3 は、碍子先端部 31 が突出するように取付金具 2 に挿嵌保持される。
	中心電極 4 の電極先端部 41 との間に火花放電ギャップ G を形成する接地電極 11	上記接地電極 11 は、図 2 に示すごとく、電極先端部 41 との間に火花放電ギャップ G を形成する。
	中心電極 4 の直径は、1.2～2.2mm	また、上記碍子固定部 22 の軸方向位置における中心電極 4 の直径は、例えば、1.2～2.2mm とすることができる。

2.1.3. 過去の特許庁調査における翻訳品質評価手法の内容

本節では、特許庁が過去に実施した下記の各調査において使用された翻訳品質評価手法の内容についてまとめる。

- ・平成 20 年度「中国語機械翻訳技術に関する調査」
- ・平成 22 年度「特許庁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」
- ・平成 23 年度「中国特許文献の和文抄録作成に対する機械翻訳の活用に関する調査」

2.1.3.1. 平成 20 年度「中国語機械翻訳技術に関する調査」[21]

中日機械翻訳の品質向上のための方策に関する調査であり、特定の中日機械翻訳エンジンに対して実施したユーザ辞書や定型利用翻訳によるチューニングの効果を測るため、チューニング前後の機械翻訳結果の比較評価が実施された。

評価は調査の過程で二度行われたが、より詳細な二度目の評価の内容は以下のとおりである。

[評価対象]

中国特許文献 6 文献の全文（600 文）を対象に、同一の中日機械翻訳エンジンのチューニング前後の出力結果を比較した（ただし下記「正確さ」の評価のみは 6 文献中 2 文献のみ対象）。

[評価方法及び基準]

「総合評価」と、「補足評価」として「正確さ」及び「流暢さ」の観点からの評価が行われた（全て 5 段階評価）。評価は、中国語が堪能な日本人 1 名と日本語が堪能な中国人 1 名、計 2 名が担当し、文献中の全ての文を文単位で評価したうえで、その平均値が求められた。

①総合評価

5 段階評価で実施された総合評価の評点基準は下表 2-15 のとおりである。翻訳者が人手翻訳をする際の素材（下訳）としてどの程度有用かを評価しており、「仮設手法」（添付資料 1-2）の「3. 絶対評価手法」に相当する、特定の用途への有用性を測る調査といえる。

表 2-15. 総合評価の評点基準

評点	説 明
5	訳文がそのまま翻訳結果として使える
4	訳文を基本にして修正すれば翻訳結果として使える
3	部分的に機械翻訳結果が利用できるが訳文の再構成が必要
2	一から自分で翻訳し直した方が効率的
1	翻訳結果に使えるところがほとんどない

②補足評価

補足評価は「正確さ」と「流暢さ」から成る。それぞれの 5 段階評価基準は下表 2-16 及び表 2-17 のとおり。

表 2-16. 正確さの評点基準

評点	説 明
5	原文の意味を過不足なく読み取ることができる
4	多少不適切な単語が混じるが全体として理解可能
3	文の趣旨は読み取れる
2	断片的に情報があるものの、文の趣旨が読み取れない
1	意味不明、または、得られる情報がない

表 2-17. 流暢さの評点基準

評点	説 明
5	ネイティブレベルの申し分のない日本語
4	多少の不自然さはあるがスラスラ読める
3	読解に時間がかかるが何とか読み進められる
2	部分的に訳されているものの全体として文を成していない
1	翻訳失敗、または、単語の羅列

2.1.3.2. 平成 22 年度「特許庁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」[22]

特許庁が海外特許庁に対し「高度産業財産ネットワーク（AIPN）」を通じて提供している特許文献日英機械翻訳結果と、これと比較するために同じ文献を市販の翻訳ソフトウェアで翻訳した結果に対し、人手評価が実施された。

[評価対象]

日本の公開特許公報 206 文献、拒絶理由通知書 305 文献、計 511 文献より各 1 文ずつ、合計 511 文が評価された。

[評価方法及び基準]

電気・物理・化学分野別に専門知識を有する評価者を配し、「意味理解度評価」と「構文理解度・訳語評価」の 2 種類の評価が実施された。以下にその基準を示す。

①意味理解度評価

原文の意味がどの程度機械翻訳文に正しく反映されているかを 5 段階評価し、全ての文の平均値を求めて品質評価結果とした。評価基準は表 2-18 のとおり。

表 2-18. 意味理解度評価の評価基準

評点	説 明
4 点	原文の意味を正しく理解できる
3 点	一部不適切な訳語があるが、概ね理解できる
2 点	構文の取り違い等があり、部分的な理解にとどまる
1 点	理解不能
0 点	翻訳失敗や訳漏れ等

なお平均値の算出方法としては、単純に全ての文の評価値を対象とする方法と、実際に人間が理解できるもののみを評価対象とし、4 点と 3 点（ただし×0.7 とし実際 2.1 点）のみを加点対象とし、2 点以下の場合は 0 点とみなす方法の 2 種類が採用されている。

②構文理解度・訳語評価

原文の構文を正しく把握できているか（構文理解度）、および適切な訳語を用いて訳出されているか（訳語評価）が、それぞれ表 2-19、表 2-20 に示す分類基準に沿ってチェックされ、該当する翻訳ミス数がカウントされた。その後、カウント値から文あたりの誤り数と文節あたりの誤り率を算出し、双方の機械翻訳結果を比較している。

表 2-19. 構文理解度評価の分類基準

大分類	小分類	説 明
形態素解析	—	形態素解析で、単語分割、品詞付与を誤った場合。
構文解析	主語	主語の係り先誤り。
	従属節	連用性の述語文節の係り先誤り。
	用言並列	並列構造となる用言文節の並列先の解析誤り。
	名詞並列	並列構造となる名詞文節の並列先の解析誤り。
	複合並列	複数の句のペアが並列をなす場合の解析誤り。
	名詞句	名詞句を構成する文節の係り先誤り（名詞並列除く）。
	その他	上記以外。
生成	—	構文的な把握が正しくても訳文生成の際に不適切な処理をしていると思われる場合。
【】 分断		【】 で文が分断される場合。
対応不明		原文にある語が訳文に無い場合や、原文に無い語が訳文にある場合。
その他		原因がはっきりしないもの

表 2-20. 訳語評価の分類基準

分 類	説 明
一般語	特定の技術分野に属するものではない一般の単語の訳語の誤り。
一般語（複合名詞）	一般語のうち複合名詞であるもの。
専門用語	特定の技術分野に属する専門用語の訳語の誤り。
専門用語（複合名詞）	専門用語のうち複合名詞であるもの。
複合動詞	副詞や名詞などと複合化した動詞の訳語の誤り。
前置詞	前置詞の選択誤り。
未知語	未知語として出力されたもの。

※一文中で同じ単語が複数回誤って訳される場合は1回としてカウント。

2.1.3.3. 平成 23 年度「中国特許文献の和文抄録作成に対する機械翻訳の活用に関する調査」[23]

この調査における品質評価は、いくつかの手法で作成された中国公報要約の和文抄録の有用性評価として行われた。対象となった和文抄録中、2 種（中国語要約の中日翻訳と英語要約の英日翻訳）が機械翻訳を利用して作成されたものであり、他に人手翻訳（中日、英日）2 種と、中国語公報全文から直接人手で作成された和文抄録 1 種、合計 5 種が評価対象とされた。

品質評価のうち、翻訳精度に関する評価の内容は下記のとおり。

[評価対象]

40 件の中国特許文献について作成された 5 種類の和文抄録が対象（40 文献×5 種類）。評価は文単位ではなく、抄録全体に対する評価である。

[評価方法及び基準]

評価は「翻訳精度評価」と「検索精度評価」の 2 つの観点で実施された。両観点とも、その下層にいくつかの項目が設定されており、各項目について 5 段階で評価したうえで、その平均値を観点の評価値とする形式になっている。なお、評価者には実際に米国・欧州の和文抄録を作成している者が起用された。

①翻訳精度評価

各和文抄録の抄録文全体に対し、翻訳精度に影響する主な要因として下記 4 つの項目を設定し、それぞれを表 2-21 に示す共通の基準にて 5 段階評価を行なったうえで、4 項目の平均値を「翻訳精度評価」の評価値としている。

- (a) 文法的な誤訳の有無
- (b) 訳漏れ、不要な語句の有無
- (c) 意味不明又は一般的でない訳語の使用の有無
- (d) 日本語文章構成としての不自然の有無（※上記 a～c に該当しない問題点）

表 2-21. 翻訳精度評価の評点基準

評点	説 明
5	問題箇所なし (0%)
4	ほぼ問題箇所なし (～10%)
3	若干の問題箇所を含む (～30%)
2	かなりの問題箇所を含む (～50%)
1	多くの問題箇所を含む (51%～)

②検索精度評価

「検索精度評価」は、評価者が各中国特許文献の内容を踏まえて選定した 10 語前後の技術用語を対象とし、その適訳性を表 2-22 に示す基準に則り 5 段階評価している。

表 2-22. 検索精度評価の評点基準

評点	説 明
5	選定キーワードの全ての訳語が的確である。
4	選定キーワードの 15%以下の訳語が不的確である。
3	選定キーワードの 30%以下の訳語が不的確である。
2	選定キーワードの 45%以下の訳語が不的確である。
1	選定キーワードの 45%以上の訳語が不的確である。

2.1.4. 本調査に有用な知見

本節では、ここまでにまとめた既存の主要な機械翻訳結果の翻訳品質評価手法についての調査内容から、本調査に特に有用と思われる知見について整理・考察した。

2.1.5.1. Adequacy と Fluency

2.1.1.にて概観した過去の主要な人手評価手法の調査結果からは、Adequacy に代表される「機械翻訳結果が原文の内容を正確に反映しているか否か」という評価観点と、Fluency に代表される「機械翻訳結果単体での理解容易性、流暢さ」を測る評価観点とが併用され、それぞれを 5 段階評価するという手法が多用されていることがわかる。

しかしながら、本調査にて作成する品質評価手法は特許文献の機械翻訳結果の実務への有用性を評価するためのものであることを考えると、「原文の内容を正確に反映しているか否か（≡Adequacy）」という観点が最も重要であることは言うまでもない。それが「理解しやすく、流暢であるか（≡Fluency）」という観点は、あくまで内容の正確さが同等である場合に考慮すべき事項であろう。

事実、2.1.2.で詳述した NTCIR-9/NTCIR-10 特許機械翻訳タスクにおける人手評価では、原文の内容を反映していなくても流暢な文であれば高いスコアが付与される、という Fluency の課題を踏まえ、Acceptability という新たな基準を設けて、原文の内容を正確に理解できない文は、たとえ流暢であっても最低評価としている。ただし、Acceptability は、誤訳が文の一部のみでも文全体でも区別なく最低評価となるため、中／低品質な翻訳結果に対する評価の解像度が低いという別の課題がある。現状の特許文献の機械翻訳品質レベルや、特許文献の内容理解においては必ずしも完璧な翻訳でなくてもある程度の有用性が見込める点を考えると、Acceptability は評価基準としてやや厳しすぎる懸念がある。

したがって、特許文献の機械翻訳結果の有用性の評価手法としては、内容の正確さをメイン評価に据え、流暢さについては補助的な評価観点として別途評価し、メイン評価で優劣がつかない場合に参照するという形が適切であろうと思われる。

なお、「内容の正確さ」「流暢さ」は、共に主観的な評価であるため、評価基準が曖昧で信頼性が低いという指摘もある。この点への対応としては、複数名の評価者による同一文の重複評価や、NTCIR で行われた基準整合のための評価者の事前意識合わせ（2.1.2.4.）、平成 23 年度特許庁調査等で導入された評価基準へのパーセンテージ併記（2.1.3.3）等が有効であろう。

2.1.5.2. 特定用途への有用性の評価

今回調査した品質評価手法には、NTCIR-10 で実施された特許審査評価（2.1.2.7）をはじめ、Mu プロジェクトにおける修正容易度・誤解釈度（2.1.1.2）、平成 20 年度特許庁調査における、人手翻訳の下訳としての有用性の評価（2.1.3.1）など、特定の用途を想定して

その有用性を直接評価する形式の評価も多数実施されている。いずれも、機械翻訳結果の用途を明確にすることで、例えば部分的に訳せている文がどの程度有用であるかを明確に判定することができ、評価対象となった機械翻訳システムが実務においてどの程度役に立つレベルであるかをより正確に評価することができている。

ただし、こうした評価は、特定の用途にカスタマイズされた形で行われるため、評価結果を他の用途に応用しにくく、特許文献のように機械翻訳結果の用途が多岐にわたる場合、用途ごとに別々の手法で評価を行う必要が生じる。本調査にて作成する「特許庁手法」は、複数の用途への有用性評価に用いられる前提であることから、特定の用途に特化した評価手法とすることは難しいが、例えば各評価観点をモジュール化し、評価値を個別に算出・重み付けできるような形式にする等、多様な用途に対応できるよう配慮することが求められる。

評価観点のモジュール化の参考とすべき事例としては、平成 23 年度特許庁調査(2.1.3.3.)にて実施された検索用途の観点からの品質評価が挙げられる。この手法は、翻訳対象文献に含まれる重要なキーワードを 1 文献あたり 10 語程度あらかじめ選定しておき、このうち適切に翻訳された語の割合で評価を行うというものである。したがってここでは係り受けの正確さ等は一切考慮されないが、検索にヒットするか否かという観点からは、キーワードとして用いられる主要な技術用語の訳質が最も重要であり、この要素のみを単独の評価観点としてモジュール化することは有効と考えられる。ただし、実際の検索においては語の順序等も配慮される場合があるので、キーワードの翻訳精度のみを検索用途の評価基準としている点には改善の余地がある。各評価観点をモジュール化することで、検索用途への有用性の評価においては、キーワードの翻訳精度を主要な評価値としつつも、内容の正確さ等の観点も加味する、といった調整が可能となる。

2.1.5.3 客観的評価の導入

Adequacy や Fluency といった主観的な評価に代えて、より客観的な観点による品質評価手法が用いられる場合もある。主な手法としては、機械翻訳結果に含まれる全ての翻訳ミスを抽出・分類し、その個数を比較する手法(2.1.1.8.)等が挙げられる。

しかしながら、この評価手法は、原語(原文の言語)と目的言語の双方を理解できる評価者でないと正確な作業が困難であるため、主観的評価に比べて評価者の確保が難しい。また、翻訳結果から全ての翻訳ミスを抽出し分類するという作業は極めて難易度が高く、評価における作業負担が過大となる懸念がある。加えて、個々の翻訳ミスが内容の正確さに与える影響はその度合いや発生箇所等により様々であるため、翻訳ミスの多寡が特定の用途への有用性の多寡と常に正確に比例するとは限らない。

これらの理由により、特許文献の機械翻訳結果の実務への有用性を評価する場合に、こうした形式の評価手法を導入することは効率的とはいえず、また必ずしも評価精度を高め

るともいえない。主観的評価に対する客観性の確保は、複数名での重複評価等、運用方法の工夫にて達成を図るほうが効率的であると考え。また、前述の「技術用語の翻訳精度」という評価観点、技術知識を有する者であれば評価の揺れは少ないと思われるため、これを一観点として組み入れることで評価の客観性を高めることも期待できる。

なお、本調査にて作成する「特許庁手法」の想定する評価結果の使途のひとつに、評価した機械翻訳システムの性能改善のためのフィードバック用途が挙げられているが、この用途に関しては、主観的評価による 5 段階評価の評点では、そのシステムにおいてどのような翻訳ミスが多発しているか等の情報を直接得ることができないため、上記したような翻訳ミスの抽出・分類作業が必須となる。したがって「特許庁手法」には、フィードバック用途に限定した形でこの形式の評価手法も含めておく必要がある。

2.1.5.4. 評価対象文数

評価対象文数は、多ければ多いほど評価の信頼性が向上していくことは言うまでもない。しかしながら、人手評価は作業コスト（時間、労力）が大きく、費用対効果の観点からは、一定の信頼性を確保したうえで極力少量に抑えることが望ましい。

今回の調査で判明した各品質評価手法における評価対象文数を列举すると、ALPAC Report (2.1.1.1.) が 144 文、Mu プロジェクト (2.1.1.2.) が 1,682 文、WMT (2.1.1.4.) が 3,000 文、NTCIR-9/NTCIR-10 (2.1.2.4.) が 300 文、平成 20 年度特許庁調査 (2.1.3.1.) が 600 文、平成 22 年度特許調査 (2.1.3.2.) が 511 文と、大きなばらつきがあり、どの程度の文数が適切かをこれらから判断することは難しい。

ただし、このうち NTCIR-9 では、評価に使用した評価対象文数 300 文の妥当性の検証のため、これを 150 文ずつに分割し、双方における Adequacy 評価に基づくシステム間の相対評価値を比較する分析が行われている (2.1.2.6.)。その結果、双方の相対評価値は高い相関を示し、システム間の評価の差が大きい場合、システム間の比較結果の信頼性は 150 文においても高いことが確認された。

NTCIR-9 は特許文献からランダムに抽出された文を対象とした品質評価であり、本調査の品質評価手法と条件に近い。このため、本調査においても、最低 150 文程度、可能であれば 300 文程度を評価対象とすることが望まれる。

ただし、複数名による重複評価を導入する場合、のべの評価実施文数は評価対象文数×評価者人数となるので、これを踏まえた対象文数の設定が必要である。

2.2. 英日機械翻訳の内容

本項では、民間企業や研究機関等にて開発された主要な英日機械翻訳エンジンについて、その概要を調査した。以下、その調査結果を記す。

2.2.1. 市販ソフトウェア（国内）

現在市販されている主な英日機械翻訳ソフトウェアを以下に挙げる。

2.2.1.1. The 翻訳プロフェッショナル V15 特許エディション [24]

東芝ソリューション株式会社「The 翻訳シリーズ」最上位製品の特許文献対応版である。英日／日英翻訳用で、エンジンはルールベース方式（RBMT）を採用している。

基本辞書に加え、31 種・計 312.2 万語の英日専門用語辞書及び 21 万語の特許用辞書（日英との合計）を備えており、最大 16 種まで同時に使用可能である。また、ユーザ自身によるユーザ辞書（こちらも最大 16 件まで同時使用可能）、パターン辞書（変数を含む文パターン登録）、翻訳メモリ（頻出文を登録）を作成・利用可能であり、学習辞書（備付辞書中の訳語間における優先語の指定）、マスク辞書（見出し語の無効化）等も使用できる。

特許文献対応としては、先述の特許用辞書のほか、特許定型表現への対応や、特許に頻出の長文を適切な位置で区切って翻訳する自動前編集機能等を備えている。

- ・方式：ルールベース方式（RBMT）
- ・販売：東芝ソリューション株式会社
- ・価格：192,150 円（税込・希望小売価格）

[公式の翻訳例]

原 文	A translation apparatus according to claim 1, wherein said component unit of at least one of said first language text and said second language text is a paragraph or sentence.
訳 文	前記第一言語テキストおよび前記第二言語テキストの少なくとも 1 つの前記構成要素のユニットはパラグラフまたは文であることを特徴とする請求項 1 に記載の翻訳装置。

2.2.1.2. PAT-TranserV11 プロフェッショナル [25]

株式会社クロスランゲージ製の特許用英日／日英翻訳ソフト「Pat-Transer」の最新・最上位版。エンジンはルールベース方式（RBMT）である。

英日辞書総数は 348.6 万語であり、38 種類の専門用語辞書と特許用基本語辞書（127.5 万語）を含む。指定可能辞書数は、ユーザ辞書を含めて 10 種。ユーザ辞書、翻訳メモリ、学習辞書等が使用できる。

特許明細書の翻訳に特化しており、特許文翻訳専用にチューニングされた翻訳辞書や、特許用例 2 万例の翻訳メモリ、さらに特許文の構造を解析し、適切な位置での長文の自動分割や、定型的表現への対応等を行う「特許文前処理・後処理機能」等を備える。

- ・方式：ルールベース方式（RBMT）
- ・販売：株式会社クロスランゲージ
- ・価格：176,400 円（税込・標準価格）

[公式の翻訳例]

原 文	I claim: A method of operating a computer system having a processor, a memory, a nonvolatile storage, text entry means, pointer positioning means, execute key means, display means and means for displaying multiple windows, comprising the steps of:
訳 文	[特許請求の範囲] 以下を有するコンピュータ・システムを作動する方法： プロセッサ、メモリ、不揮発性記憶装置、テキスト入力手段、ポインタ位置決め手段、実行キー手段、表示手段、そして、複数のウインドウを表示するための手段；上記は、次のステップを備える：

2.2.1.3. ATLAS V14 翻訳スーパーパック [26]

富士通株式会社による英日・日英翻訳ソフトの最上位モデルであり、エンジンはルールベース方式（RBMT）を採用している。

英日基本辞書 143 万語と、28 分野 557 万語の専門用語辞書（ただし語数は日英との合計）を利用可能。専門用語辞書はユーザ辞書と合わせて最大 32 種まで同時指定可能である。ユーザ辞書、翻訳メモリも利用可能で、学習辞書に相当する機能もある。

先の 2 製品と異なり、特許文献への対応は謳われておらず、特許用辞書も装備していないが、翻訳メモリとして化学特許分野 15 万語、電気特許分野 20 万語が収録されている。また、特許特有の文に対する解釈を優先させる設定（特許文用文法）が可能である。

- ・方式：ルールベース方式（RBMT）
- ・販売：富士通ミドルウェア株式会社
- ・価格：134,400 円（税込）

[公式の翻訳例] ※マニュアルより上記「特許文用文法」の例を使用。

原 文	means to spin said device for 60 s
訳 文	60 秒間前記の装置を回転させる手段

2.2.1.4. その他の市販ソフト

上記 3 社製品以外で、著名通販サイトにて新品の取扱いがあるものについて調査した。同一メーカーの製品が複数種該当する場合は、最新・最上位製品と思われるものを調査対象とした。なお販売価格が 10,000 円未満のものは、ヘビーユース向きでないとみなし、対象外とした。

[英日・日英翻訳ソフト翻訳 J・E・T PRO 版]

基本辞書 227 万語に加え 142 万語の専門用語辞書を収録した英日・日英翻訳ソフト。株式会社高電社製、調査時の販売価格 25,200 円。2004 年発売と古く、同社ウェブサイト上

にも掲載されていないことから、本シリーズは販売終了となった可能性がある。高電社は現在も多数の翻訳ソフトウェアを取扱っているが、大半は中国語、韓国語関連であり、英語の機械翻訳はインターネット経由のASP型翻訳サービス「J-Server オフィスパーソナル」のみ。WORD等のオフィスソフトにアドインを組み込んで使用する仕組みであり、特許文献の大量翻訳向きの製品とはいえない。

[LogoVista Pro2014 フルパック] [27]

英日辞書 552 万語(うち専門用語辞書は 34 分野 243 万語)を搭載した英日／日英ソフト。ロゴヴィスタ株式会社製、税込価格 89,800 円。

特許用辞書はないが、専門用語辞書にて主要な技術分野はカバーされている。特許関連では、「技術文献検索機能」(調べたい単語を日本語入力するだけで、単語を辞書引きしてから Web 上で公開されている技術文献データベースの中から目的の文献を検索する機能)の中で、アクセスできる Web データベースとして USPTO、EPO が挙げられているが、翻訳支援機能に属するものであり、特許文書の機械翻訳機能ではない。

[Macintosh 用翻訳ソフト EtoJ PERSONAL Ver.4.5 英日翻訳]

Macintosh 用の英日翻訳ソフト。カテナ社(吸収合併を経て現在は株式会社システナ)製品だが、同社ウェブサイトには本製品を含む翻訳ソフトの紹介は見当たらない。調査時の販売価格は 41,790 円。

[翻訳ブレイン 3] [28]

株式会社ジャストシステム製の英日／日英翻訳ソフト。2011 年発売、調査時の販売価格は 36,540 円。辞書語数は基本語 118 万語、専門辞書 22 種 112 万語の計 230 万語。カタログ中に特許に関する言及は見当たらない。

[本格翻訳 8 Platinum] [29]

ソースネクスト株式会社製の日英／英日翻訳ソフト。本製品はシリーズ最新版だが、ウェブ限定販売。カタログには、基本語 338 万語、専門用語辞書 106 分野 360 万語と記載されているが、英日単独での語数は不明。特許辞書は装備していない。

2.2.2. 市販ソフトウェア（海外）

海外の市販ソフトウェアのうち特に著名なもの 2 種について調査した。

2.2.2.1. SYSTRAN 7 Premium Translator

仏・シストラン社の販売する RBMT 方式の機械翻訳ソフトウェアの最新版。英語と、日本語を含む 12 ヶ国語間の翻訳に対応。分野別辞書(Domain Dictionaries)は全 20 種類とあるが、収録語数や対象言語等は明示されておらず、詳細不明であった。

2.2.2.2. TRADOS [30]

トラドス社（現 SDL インターナショナル社）が開発した翻訳支援ツールであり、蓄積した過去の対訳文から翻訳メモリを利用して類似の文を表示するもの。特許翻訳時に利用されることも多いが、翻訳を完成させるのは人間であり、全自動で訳文を出力する機械翻訳ソフトとは異なる。

2.2.3. ウェブ上で利用可能な無料翻訳サービス

インターネット上で利用可能な主な無料英日翻訳サービスを以下に示す。

2.2.3.1. 大手プロバイダによるサービス

Google、Yahoo!、Excite、@nifty 等のプロバイダが無料で利用できる機械翻訳サービスを提供している。ウェブ上でのサービス形態は、翻訳したい文をボックスに貼りつけるテキストボックス方式と、翻訳したいページの URL を指定するウェブ方式が中心である。

Google 翻訳には、Google 自らが開発した統計方式（SMT）の機械翻訳エンジンが用いられている。調査時点では 71 ヶ国語間の翻訳に対応している。

Yahoo!翻訳は 9 ヶ国語に対応。エンジンはクロスランゲージ社製のものを使用しており、方式は非公開だが英日翻訳に関してはルールベース方式（RBMT）と考えられる。

Excite 翻訳は、高電社製のエンジンを使用しており、日本語と、英語を含む 8 ヶ国語間での翻訳が可能。

@nifty 翻訳では日本語と英中韓 3 ヶ国語間の翻訳が可能。エンジンは英日間は富士通 ATLAS を使用している。

2.2.3.2. EPO Patent Translate

EPO と Google が共同で開発した特許文献用の機械翻訳システム。EPO サイトには「特許文献の語彙や文法に対応するために特別の学習を施した機械翻訳システム」とある。具体的には、Google の保有する SMT 方式の機械翻訳エンジンに、EPO の提供する特許文の対訳データを学習させることで、過去の特許文で使われる用語や言い回しが翻訳結果に反映されるようになっている。

EPO の Espacenet で検索した文献に対し、要約、クレーム、明細書本文の単位で翻訳が可能。タイトルは対象外である。

ただし、英日翻訳に関しては、欧米の特許文献 8 文献を比較した結果、PATENT TRANSLATE での翻訳結果と、Google 翻訳での翻訳結果との間に有意な違いは見出せなか

った。このことから、少なくとも英日翻訳に関しては、両者は同等の環境（エンジン、学習コーパス）を使用しているものと考えられる。

2.2.3.3. 訳してねっと [31]

沖電気工業株式会社が開設したコミュニティ型の無料機械翻訳サイト。ユーザが協力して辞書を充実させ、翻訳品質を向上させている。英日間と中日間の翻訳に対応しており、英日エンジンは自社製のパターン翻訳方式（＝用例ベース）エンジンが使用されている。ジャンル中に「法律・特許」が存在する。

2.2.4. 研究機関の英日機械翻訳エンジン

NTCIR の特許翻訳タスク（2.1.2.参照）には、多数の研究機関が自らの機械翻訳エンジンにて参加している。参加者は、NTCIR より提供される特許文献対訳データを学習用データとして使用することができる。

2012～2013 年度に行われた NTCIR-10 の英日特許翻訳タスクには、11 機関・団体による 18 のエントリがあった（主催者自身によるベースライン用エントリを除く）。このうち 14 エントリ（7 機関）を SMT 方式が占め、それ以外は RBMT 方式が 1 エントリ（1 機関）、用例ベース方式（EBMT）が 2 エントリ（2 機関）、そしてハイブリッド方式が 1 エントリ（1 機関）であった。

各エントリは人手評価（Adequacy、Acceptability）と自動評価（RIBES、BLEU、NIST）により品質を評価された。このうち Acceptability の評価では、第 1 位と第 2 位を NTT コミュニケーション科学基礎研究所のエントリ（SMT）が占め、以下、第 3 位が山梨英和大学のエントリ（ハイブリッド）、第 4 位が一般財団法人日本特許情報機構のエントリ（RBMT）という順位となった。以下、第 1 位から第 4 位までのエントリで使用された英日機械翻訳システムについて記す。

2.2.4.1. NTT コミュニケーション科学基礎研究所

英語原文を日本語に翻訳しやすい語順に予め並べ替えて（Preorder）から英日翻訳することを特徴とする SMT 方式の英日機械翻訳システム。語順並べ替えには日本語の主辞後置性に着目したルールを用いている。後段の翻訳にはフレーズベースの Moses を使用している。語順の並べ替えに必要な構文解析器として、特許分野に適応した依存構造解析器（dependency parser）を活用している。本依存構造解析器は、半教師あり学習法により分野適応を行っている。事前並べ替えに基づくシステムと、それを他の 2 種類のシステム（Moses-Chart と Moses）と組み合わせたコンビネーション・システムの 2 種類の翻訳結果が提出された。

2.2.4.2. 山梨英和大学

市販の RBMT 方式ソフトウェアを用いて英語を日本語に翻訳したのち、統計的自動後編集 (SPE: Statistical Post Editing) を施して、より良い日本語にするというハイブリッド方式である。SPE の処理系としては、フレーズベースの Moses を使用している。

2.2.4.3. 一般財団法人日本特許情報機構 (Japio)

市販の RBMT 方式ソフトウェアに対して特許文献用の各種チューニングを施し、さらに各種特許文献から独自に収集・編纂した分類別の特許技術用語辞書 (34 分類、平均 6.9 万語) を追加したシステムである。

2.2.5. 本調査に有用な知見

現在、国内で市販されている主要な英日機械翻訳ソフトウェアはすべて RBMT であり、無料のウェブサービスも、これら RBMT エンジンのいずれかを組み込んだものが多い。しかしその一方で、Google 翻訳やこれを特許文献に適用した EPO の Patent Translate では SMT エンジンが採用されており、NTCIR-10 の特許機械翻訳タスクでも、最も優秀な評価を得た NTT コミュニケーション科学基礎研究所のエンジンが SMT、第 3 位の山梨英和大学のエンジンはハイブリッド方式と、RBMT 以外の方式も実用化が進んでいる。

「特許庁手法」は、こうした状況を踏まえ、多様な方式の機械翻訳結果を公平に評価できるような評価基準を定める必要がある。

【参考文献】

- [1] Automatic Language Processing Advisory Committee, National Research Council: Language and Machines, Computers in Translation and Linguistics, 1966.
- [2] 長尾真、辻井潤一: Mu プロジェクトにおける日英翻訳結果の評価、情報処理学会研究報告、自然言語処理 47-11、pp.79-88、1985。
- [3] 長尾真: 機械翻訳文の質の評価と言語の制限、情報処理 Vol.26、 No.10、 pp. 1197- 1202、1985。
- [4] LDC: Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Arabic-English and Chinese-English Translations, 2002.
- [5] Philipp Koehn and Christof Monz: Manual and Automatic Evaluation of Machine Translation between European Languages, Proceedings on the Workshop on Statistical Machine Translation, pp.102-121, 2006.
- [6] Chris Callison-Burch, Cameron Fordyce, Philipp Koehn, Christof Monz and Josh Schroeder: (Meta-) Evaluation of Machine Translation, Proceedings of the Second Workshop on Statistical Machine Translation, pp.136—158, 2007.
- [7] Atsushi Fujii, Masao Utiyama, Mikio Yamamoto, Takehito Utsuro: Overview of the Patent Translation Task at the NTCIR-7 Workshop, Proceedings of NTCIR-7 Workshop Meeting, pp.389-400, 2008.
- [8] Shoichi Yokoyama, Hideki Kashioka, Akira Kumano, Masaki Matsudaira, Yoshiko Shirokizawa, Shuji Kodama, Terumasa Ehara, Shinichiro Miyazawa and Yuzo Murata: An automatic evaluation method for machine translation using two-way MT, Proceedings of the MT Summit VIII, pp.379-384, 2001.
- [9] Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu: Bleu: a Method for Automatic Evaluation of Machine Translation, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp.311-318, 2002.
- [10] George Doddington: Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics, 2002.
- [11] Hiroshi Echizen-ya, Kenji Araki: Automatic Evaluation of Machine Translation based on Recursive Acquisition of an Intuitive Common Parts Continuum, Proceedings of the Eleventh Machine Translation Summit (MT SUMMIT XI), pp.151-158, 2007.
- [12] Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, Hajime Tsukada: Automatic Evaluation of Translation Quality for Distant Language Pairs, Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages 944–952, 2010.
- [13] Chris Callison-Burch, Miles Osborne and Philipp Koehn: Re-evaluating the Role of BLEU in Machine Translation Research, Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, pp.249-256, 2006.
- [14] 井佐原均、新納浩幸、山端潔、森口稔、野村浩郷: JEIDA 機械翻訳システム評価基準 (品質評価編) —英日翻訳の品質評価項目の検討と評価用コーパスの作成—、情報処理学会研究報告自然言語処理 (NL)、1993-NL-096、pp.81-88、1993。

- [15] AAMT 機械翻訳課題調査委員会：テストセットによる日英／日中機械翻訳エンジンの評価（Ⅰ）、<http://corpus.aamt.info/>.
- [16] David Vilar, Jia Xu, Luis Fernando D'Haro and Hermann Ney: Error analysis of statistical machine translation output, Proceedings of the LREC-2006: Fifth International Conference on Language Resources and Evaluation, pp.697-702, 2006.
- [17] 江原暉将、田中穂積：機械翻訳における自然言語処理、情報処理、Vol.34、No.10、pp.1266-1273、1993。
- [18] Isao Goto, Bin Lu, Ka Po Chow, Eiichiro Sumita, and Benjamin K. Tsou. 2011. Overview of the patent machine translation task at the NTCIR-9 workshop. In Proceedings of NTCIR-9, pages 559–578.
- [19] Isao Goto, Ka Po Chow, Bin Lu, Eiichiro Sumita, and Benjamin K. Tsou. 2013. Overview of the patent machine translation task at the NTCIR-10 workshop. In Proceedings of NTCIR-10, pages 260–286.
- [20] LDC. 2005. Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Translations Revision 1.5. Linguistic Data Consortium. <http://projects.ldc.upenn.edu/TIDES/Translation/TransAssess04.pdf>.
- [21] 特許庁平成 20 年度「中国語機械翻訳技術に関する調査」報告書
- [22] 特許庁平成 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」報告書
- [23] 特許庁平成 23 年度「中国特許文献の和文抄録作成に対する機械翻訳の活用に関する調査」報告書
- [24] 東芝ソリューション株式会社 製品情報ウェブサイト
http://pf.toshiba-sol.co.jp/prod/hon_yaku/premium/index_j.htm
- [25] 株式会社クロスランゲージ 製品情報ウェブサイト
http://www.crosslanguage.co.jp/products/pat-transer_v11/
- [26] 富士通株式会社 製品情報ウェブサイト (<http://software.fujitsu.com/jp/atlas/>) 及び製品マニュアル
- [27] ログヴィスタ株式会社 製品情報ウェブサイト
https://www.logovista.co.jp/LVERP/shop/ItemDetail.aspx?contents_code=LVXEFX14W
- [28] 株式会社ジャストシステム 製品情報ウェブサイト
<http://www.justsystems.com/jp/products/honyaku/>
- [29] ソースネクスト株式会社 製品情報ウェブサイト
http://www.sourcenext.com/product/pc/use/pc_use_000863/
- [30] SDL インターナショナル社 製品情報ウェブサイト
<http://www.translationzone.com/jp/products/sdl-trados-studio/>
- [31] コミュニティ型機械翻訳サイト 訳してねっと！ウェブサイト
<http://yakushite.info/index.ja.html>

第 3 章

英日機械翻訳の現状分析

第3章 英日機械翻訳の現状分析

「特許庁機械翻訳品質評価手法」の策定に先立ち、その評価対象となる英日機械翻訳の現状分析のため2種類の調査を実施し、主要な英日機械翻訳エンジンにおける重要技術用語の誤訳率の確認と、英日機械翻訳に特有の課題・留意点の洗い出しを行なった。本章にて、これらの調査の内容及び結果について記す。

3.1. 英日機械翻訳における重要技術用語の翻訳精度の評価

現状の英日機械翻訳における技術用語の翻訳精度を把握するため、第2章「文献調査」にて調査した英日機械翻訳エンジンのうち5種類を用いて、英語特許文献20文献を機械翻訳し、各文献の機械翻訳結果（20文献×5エンジン）に含まれる重要技術用語の翻訳品質を評価した。

3.1.1. 英日機械翻訳エンジン

調査に使用する英日機械翻訳エンジンとして、ルールベース方式から3種類（RBMT1～3）、統計翻訳方式から2種類（SMT4～5）を選定した。

RBMT1～3は、初期設定にて使用しており、特段のチューニング等は施していない。ただし専門辞書の設定は、IPCセクションごとに、各製品付属の専門辞書を適当と思われる順に最大限（指定可能最大数）組み込んで使用した¹。一方、SMT4～5に関しては、いずれもユーザ側での環境設定は不可能であったため、それぞれの（調査時点の）実用環境のまま使用した。

したがって、各エンジンの翻訳条件は一定ではなく、本調査結果をそのまま各エンジンの性能比較に用いることは推奨しない。

3.1.2. サンプル英語特許文献

本調査にて対象とするサンプル英語特許文献20文献は、米国公開特許文献及び欧州公開特許文献より、日本語パテントファミリーを有しない（すなわち日本語対応文献の存在しない）文献を、各IPCセクションから偏りなく選定した。なお、各サンプル文献は、評価時のリファレンスに用いるため、日本人技術翻訳者を起用して全文を和訳した。

サンプル文献20文献の詳細（IPC、文献番号、発明の名称）は、添付資料3-1『英日機械翻訳の現状分析』サンプル文献一覧」を参照のこと。

¹ 各RBMTエンジンのIPCセクション別専門辞書設定は、「辞書設定データ.xls」（電子データ）に示す。

3.1.3. 重要技術用語の選定

評価対象となる重要技術用語は、各文献から 10 語ずつ、計 200 語選定した。各用語は、各文献の和訳を担当した日本人技術翻訳者が、「文献の発明内容を理解するために最も重要な技術用語」という観点で選定した。このため、必ずしも翻訳の難易度の高い用語が選ばれているわけではない。なお、選定者が必要と認めた場合は、名詞以外の語を選択することも許容した。

下表 3-1 に、選定された重要技術用語 200 語の品詞別語数を示す。各文献における選定用語の一覧は添付資料 3-2 『英日機械翻訳の現状分析』文献物重要技術用語一覧」を参照のこと。

表 3-1. 重要技術用語 品詞別語数

品詞	用語数	例
名 詞	191	coupling assembly, pulley, electrically conductive textile yarn
動 詞	1	receive
形容詞	5	semiautogenous, autogenous, offset, grill shaped, machine washable
副 詞	2	telescopically, in situ
その他※	2	(0001) plane or (000-1) plane, staples of first element

※その文献のみで通用する用語といえるため、他の名詞と区別した。

選定された重要技術用語には、複数の語からなる複合語も多数含まれる。各用語の構成語数分布は表 3-2 のとおりである。なお、ハイフオンやスラッシュで接続された語（例：multi-lead）は 1 語としてカウントした。

表 3-2. 重要技術用語 品詞別構成語数分布

品詞	用語数	1 語	2 語	3 語	4 語	5 語
名 詞	191	42	109	31	7	2
動 詞	1	1	0	0	0	0
形容詞	4	2	2	0	0	0
副 詞	2	1	1	0	0	0
その他	2	0	0	0	1	1
総 計	200	46	112	31	8	3

3.1.4. 用語の翻訳精度の評価基準

5 種類の機械翻訳結果における重要技術用語の訳語を、以下の 4 段階の評価基準で評価した。

- 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はないが、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。

当初、技術用語の評価基準は、「仮設手法」(添付資料 1-2) の「(2) 技術用語に関する誤訳の個数」の評価基準に従い、「人手翻訳に対して技術的に同義な範囲を外れている(未知語、訳漏れを含む)」訳語を誤訳と判定するという、いわば「適訳」と「誤訳」の 2 段階の基準を使用する予定であった。

しかし、この 2 段階の評価基準では、完全な誤訳語、すなわち、全く異なる意味の訳語に訳されており、正しい意味を類推することが不可能な訳語(例えば「light : 光」を「軽い」と訳している場合等。light が双方の意味を持つ多義語であるという原語知識がない限り、「軽い」から「光」を類推することは不可能)も、技術的には同義とは言えないが意味するところはほぼ正確に把握できるような訳語(例えば「crosslinking agent : 架橋剤」を「橋架剤」と訳す等。以下、可訳語という)も、そして未知語や訳漏れも、一律で「誤訳」と評価されることとなる。

確かに「橋架剤」のような可訳語は、検索用途への有用性という観点においては、検索式作成時にこれを想起することは困難であり、したがって検索の役には立たないため、完全な誤訳語や未知語と同等の評価で妥当である。だがその一方で、同じく特許文献の重要な用途である「原文の内容理解(粗読、精読)」の観点からは、可訳語は原語の意味自体はほぼ正しく伝達しているため、むしろ適訳語に近い評価を与えるべきである。

「仮設手法」の 2 段階の評価基準は、可訳語と誤訳語とを区別することができないため、このように用途に応じて可訳語の扱いを変えることができない。本調査にて作成する「特許庁手法」が、機械翻訳結果の主な用途として検索及び粗読・精読を想定していること(4.1.2.)を考えると、いずれにせよこの評価基準は見直す必要がある。

このことを踏まえ、本評価に用いる評価基準は、「特許庁手法」に採用することを前提に、上に示したとおり「適訳語 (A)」と「誤訳語 (C)」の間に「可訳語 (B)」を加え、用途に応じて可訳語の扱いを変更できるものを用いることとした。また、「誤訳語 (C)」と「不訳語 (D)」に関しては、検索・粗読・精読いずれの用途においても有用性がなく、同等の扱いでよいが、翻訳ミスの分類上は異なる性質のものであるため、必要に応じて区別可能なよう、別の評点とした。

3.1.5. 重要技術用語の翻訳精度評価サンプル

本評価の対象となったいくつかの重要技術用語について、評価 A～D それぞれに判定された訳語の実例を下表 3-3 に示す。

表 3-3. 重要技術用語の翻訳精度評価 評価値別訳語サンプル

用 語	multilayer coating	flexible hose	telescopically
(人手翻訳)	多層被膜	可撓性ホース	入れ子状に
評価 A 適訳語	多層コーティング 多層被覆 多層皮膜	ブレーキホース フレキシブルホース たわみホース 可撓性ホース	入れ子式に
評価 B 可訳語	多層のコーティング 多層膜	たわみ管	伸縮自在に
評価 C 誤訳語	多層を覆うこと	曲げやすい／ホース	伸縮 望遠鏡で
評価 D 不訳語	－	フレキシブル Hose	telescopically に

※表中の／は、語が分断されたことを示す。

3.1.6. 評価結果の集計方法

機械翻訳においては、同一の用語であっても、前後の語との関係等により異なる訳語が出力される場合がある。このため、重要技術用語の翻訳精度の評価は、文献中に重要技術用語が出現する都度評価する形式とした。同一文献における重要用語の出現回数は、1 回～107 回²と様々であった。

その後、全 200 語の出現の都度の評価結果（のべ 3,161 語）を集計し、A～D それぞれの率を以下のとおり算出した。

- ① 適 訳 率 = 評価 A（適訳）の語数 ÷ 「重要技術用語」総出現回数
- ② 可 訳 率 = 評価 B（可訳）の語数 ÷ 「重要技術用語」総出現回数
- ③ 誤 訳 率 = 評価 C（誤訳）の語数 ÷ 「重要技術用語」総出現回数
- ④ 不 訳 率 = 評価 D（不訳）の語数 ÷ 「重要技術用語」総出現回数

² サンプル文献 1（US2007000673）の用語#6「arm：アーム」。

3.1.7. 評価結果

5 種類の機械翻訳結果における重要技術用語の翻訳精度評価の集計結果を表 3-4 に示す。
なお、表中の「語数」はのべの出現回数である。

表 3-4. 重要技術用語の翻訳精度評価 集計結果 (サンプル 20 文献)

評価値		RBMT1	RBMT2	RBMT3	SMT4	SMT5
A 適訳	語数	2,011	2,117	2,194	2,394	2,264
	%	63.6%	67.0%	69.4%	75.7%	71.6%
B 可訳	語数	621	722	674	458	487
	%	19.6%	22.8%	21.3%	14.5%	15.4%
C 誤訳	語数	493	318	285	283	276
	%	15.6%	10.1%	9.0%	9.0%	8.7%
D 未知	語数	36	4	8	26	134
	%	1.1%	0.1%	0.3%	0.8%	4.2%
総計	語数	3,161	3,161	3,161	3,161	3,161
	%	100%	100%	100%	100%	100%

3.1.7.1. 評価 A : 適訳率

評価 A の訳語とは、「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」を指す。本調査ではこうした訳語を「適訳語」と呼び、全ての重要用語訳語に占める適訳語の割合を「適訳率」とした。5 種類の機械翻訳結果における適訳率を下表 3-5 に示す。

表 3-5. 重要技術用語の翻訳精度評価 A. 適訳率

評価値		RBMT1	RBMT2	RBMT3	SMT4	SMT5
A 適訳	語数	2,011	2,117	2,194	2,394	2,264
	%	63.6%	67.0%	69.4%	75.7%	71.6%

上表のとおり、5 種のエンジンにおける適訳率は、SMT4 が最も高く 75.7%、RBMT1 が最も低く 63.6%であった。この結果から、現状の英日機械翻訳は、1 文献に含まれる技術用語のうち平均 6~7 割程度が適切に訳される水準であると考えられる³。

重要技術用語の適訳率は、機械翻訳結果の用途のうち、特に「検索」用途に大きく関係する。仮に意味がおおむね理解できる訳語（＝評価 B、可訳語）であったとしても、検索にヒットするためには、検索者が検索式作成時に思い当たる範疇の同義語である必要があるからである。

³ ただしこの水準は文献に用いられる用語の難易度によって上下する。

ただしその一方で、検索は文献単位に行なわれることを考えると、極論すれば、文献中で 1 回でも適訳に訳されていれば、少なくとも検索にはヒットするとも考えられる。この観点から、「重要技術用語 200 語のうち、少なくとも 1 回は A 評価されたもの」を再度カウントしてみた。その結果は以下のとおりである。

表 3-6. 200 語中「A.適訳」評価の訳語を含む語数

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
A 評価あり	122 語	132 語	129 語	182 語	158 語
	61.0%	66.0%	64.5%	91.0%	79.0%
A 評価なし	78 語	68 語	71 語	18 語	42 語
	39.0%	34.0%	35.5%	9.0%	21.0%

SMT4 は、200 語中 182 語（91%）について少なくとも 1 回は A 評価を得ており、他を圧倒している。2 位も SMT5（79%）であり、この「重要技術用語が文献中で 1 回は適訳に訳される」という観点においては、SMT 方式のほうが優秀な結果を挙げている。ただし、先に示した「のべ語数」での適訳率では SMT エンジンと RBMT エンジンにここまでの差はないことから、SMT エンジンのほうが、ひとつの用語に対する訳語のゆれが大きいとも考えられる。

【訳語のゆれ】

上記考察を受け、ここで各エンジンにおける「訳語のゆれ」の度合いについて検討しておく。

たとえ同じ機械翻訳エンジンでも、同一の用語を常に同じ訳語に訳すとは限らない。例えば SMT4 では、サンプル文献において重要技術用語に選定された「coupling assembly : 連結組立体⁴」に対し、「結合アセンブリ（評価 A）」「連結アセンブリ（A）」「カップリングアセンブリ（B）」「アセンブリを結合（C）」「アセンブリを連結（C）」と、5 種類の訳語が使われている。

各エンジンにおける訳語のゆれの発生の度合いは、1 つの用語に対して複数の評価点が与えられたケースの多寡から推定できる。下表 3-7 に、重要技術用語 200 語中、出現回数が 2 回以上の語 180 語を対象に、複数の評価点が与えられた語の数を示す⁵。

表 3-7. 複数の評価点が与えられた重要技術用語数

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
2 種以上の評価値が付された語	64 語	58 語	47 語	88 語	85 語
	35.6%	32.2%	26.1%	48.8%	47.2%

⁴ サンプル文献 1（US2007000673）の用語#1。添付資料 3-1、3-2 参照。

⁵ この条件は、「複数の訳語が同一の評価値を与えられたケース」はカウント対象外となるため、実際の訳語のゆれは上記数値よりも多い可能性がある。

上記結果を見ると、やはり SMT エンジンは RBMT エンジンに比べて訳語のゆれが大きいことが分かる。SMT エンジン 2 種においては、ともに 45%以上の語において訳語のゆれが生じている。ただし、RBMT エンジン 3 種にも 25～35%程度と少なからぬ訳語のゆれが生じており、訳語のゆれは必ずしも SMT 特有の課題ではないことも判明した。

なお、上記「2 種類以上の評価値が付された語」という条件には、「おおむね同一の訳語で訳されているが、一部のみ特殊な事情（品詞誤り、解析エラー等）により異なる訳語に訳されたケース」と、「複数の訳語がともに一定以上の頻度で併用されたケース」とが混在する。両者の区別は、例えば「複数の評価値のうちの 1 つが 8 割以上を占めるか否か」といった条件で、ある程度は可能である。表 3-8 に、各エンジンの「訳語のゆれ」をこの条件で分別した結果を示す。

表 3-8. 複数の評価値が付された語中、特定の評価値が 8 割以上を占めた語

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
複数評価値の語	64 語	58 語	47 語	88 語	85 語
8 割以上あり	41 語	31 語	37 語	47 語	52 語
8 割以上なし	23 語	27 語	10 語	41 語	33 語
(180 語中の割合)	12.8%	15.0%	5.5%	22.8%	18.3%

仮に「8 割以上なし」を実質的な「訳語のゆれ」と見なすとすれば、全てのエンジンにおいてその数は「2 種以上の評価値が付された語」のうちの半数以下となる。ただし、RBMT に比べて SMT のほうが訳語のゆれが大きいという結果自体には変化はなかった。

訳語のゆれは、文献の検索においては、検索の確率を向上させるメリットがあるが、その反面、検索ノイズが多くなるというデメリットもある。また、同じ用語が文中異なる訳語で出現することとなるため、文献の粗読・精読においては読解の妨げとなる。

3.1.7.2. 評価 B : 可訳率

続いて、評価 B について分析する。評価 B は「技術用語として一般的に用いられる訳語ではないが、意味はおおむね正しい」訳語であり、本調査においてはこれを「可訳語」と呼ぶ。重要技術用語の訳語全件に占める各エンジンの可訳率は表 3-9 のとおりである。

表 3-9. 重要技術用語の翻訳精度評価 B. 可訳率

評価値		RBMT1	RBMT2	RBMT3	SMT4	SMT5
B 可訳	語数	621	722	674	458	487
	%	19.6%	22.8%	21.3%	14.5%	15.4%

この数値のみを見ると、RBMT エンジンがおおむね 20%前後、SMT エンジンが 15%前後であり、RBMT のほうが数値が高い。ただし、適訳語の割合が高い SMT エンジンにおいて可訳語の割合が低くなるのは必然であり、可訳率の正当な評価は、「全訳語のうち適訳語を除いた語に占める可訳語の割合」にて行う必要がある。次表 3-10 に、その割合を示した。

表 3-10. 適訳語を除いた語における可訳率の割合

評価値	RBMT1	RBMT2	RBMT3	SMT4	SMT5
評価 A 以外の語数	1,150	1,044	967	767	897
うち評価 B 語数	621	722	674	458	487
%	54.0%	69.2%	69.7%	59.7%	54.3%

例外はあるものの、こちらも概ね RBMT のほうが SMT よりも 10%程度高い数値を示している。ただしこれも、SMT では適訳語に訳された用語、すなわち翻訳難易度が比較的低い用語が、RBMT にはより多く残されていた可能性が高いことを考えると、これをもって RBMT が SMT より優秀であるとはいえない。

特許文献の機械翻訳結果の主要な用途のひとつとして、発明の大まかな内容を把握するための「粗読」用途や、より詳細な内容を把握するための「精読」用途が挙げられる。これらの用途においては、重要技術用語は、適訳語に訳されるのが最良であるにせよ、最低限、可訳語に訳されていれば、（読解の手間を別とすれば）同等の内容把握は可能である。したがって、「粗読」用途や「精読」用途への適否を測る場合、適訳率と可訳率を合算した率にて評価を行う必要がある。表 3-11 に、各エンジンにおける適訳率と可訳率の合算結果を示す。

表 3-11. 重要技術用語の翻訳精度評価 A. 適訳率+B. 可訳率

評価値	RBMT1	RBMT2	RBMT3	SMT4	SMT5
適訳率	63.6%	67.0%	69.4%	75.7%	71.6%
可訳率	19.6%	22.8%	21.3%	14.5%	15.4%
合算値	83.2%	89.8%	90.7%	90.2%	87.0%

上記合算結果をそのまま評価値とした場合、RBMT3 が SMT4 をわずかながら上回り、また RBMT2 も SMT5 を上回る結果となる。加えて、全てのエンジンが約 80～90%の範囲に密集していることから、現状の英日機械翻訳では、SMT と RBMT のどちらの方式のエンジンでも、標準的な難易度の文献であれば、文献中の 8～9 割の技術用語は正しい意味に訳されることが期待できるとわかった。

ただし、「粗読」用途や「精読」用途への有用性を評価する際に、適訳語と可訳語と同等に扱うべきであるか否かについては、慎重に判断すべきである。合算値が同等の場合に適訳率の高いものを上位とすべきことは言うまでもないが、「適訳率では劣るが合算値では上回る」といったケースをどのように評価するかは、機械翻訳結果の用途や具体的な使用環境等を考慮して適切に判断する必要がある。

3.1.7.3. 評価 C～D：誤訳率と不訳率

評価 C すなわち「誤訳語」は「間違った意味に翻訳された語や意味不明の語」を指し、一方、評価 D「不訳語」は「未知語：翻訳できずに原語（またはその一部）がそのまま出力されている語」及び「訳漏れ：原語に対応する訳語が翻訳結果から欠落している場合」を指す。表 3-12 に、各翻訳エンジンの誤訳率と不訳率を示す。

表 3-12. 重要技術用語の翻訳精度評価 C. 誤訳率/D. 不訳率

評価値		RBMT1	RBMT2	RBMT3	SMT4	SMT5
C 誤訳	語数	493	318	285	283	276
	%	15.6%	10.1%	9.0%	9.0%	8.7%
D 不訳	語数	36	4	8	26	134
	%	1.1%	0.1%	0.3%	0.8%	4.2%

誤訳率は RBMT1 が他のエンジンより高いことが目立つが、他の RBMT と SMT はほぼ同等であり、方式間に顕著な傾向の違いは見られない。一方、不訳率については、SMT のほうがやや高くなるという大まかな傾向が見て取れる⁶。ただしこちらも、SMT5 が際立って高く、また RBMT1 は SMT4 と同程度であり、やはりエンジンごとの傾向のほうが顕著である。

⁶ RBMT1 の未知語には、文全体が原文のまま出力された 1 文に含まれていた重要技術用語のべ 9 語を含む。これらは本来は、システムエラーとして扱うべき事象である。

なお、今回使用したサンプル文献には、OCR を利用してテキスト化されたと思われるものも含まれており、前後の語と区切られていない（文献 19 の 2 行目：withsolder bump）、語中に余分なスペースが含まれる（14 の 9 行目：re-Δfrigerant）等、未知語扱いされて当然なものもあった。このため、今回の調査にて算出された不訳率は、エンジン本来の性能よりも若干高めになっていることは否めない。本来、こうした用語は集計値から除外すべきであるが、こうした語であっても SMT エンジンでは未知語とならない場合もある（SMT4 は後者例を適切に訳していた）ことが判明したため、あえて除外せず、集計に加えている。

誤訳語と不訳語は、どちらも「検索」用途においては検索モレの原因となり、「粗読」・「精読」用途においては正しい意味の伝達を妨げるため、品質評価にあたっては、同等に扱うのが妥当である。表 3-13 に、各エンジンにおける誤訳率と不訳率の合算値を示す。この数値は、当然ながら、100%から適訳率と可訳率の合算値を減じた値と等しい。

表 3-13. 重要技術用語の翻訳精度評価 C. 誤訳率+D. 不訳率の合算値

評価値		RBMT1	RBMT2	RBMT3	SMT4	SMT5
C&D	語数	529	322	293	309	410
合算値	%	16.7%	10.2%	9.3%	9.8%	12.9%

3.1.7.4. RBMT 方式と SMT 方式の比較

今回の調査結果からは、技術用語の翻訳精度という観点に限れば、SMT 方式のエンジン 2 種のほうが RBMT 方式のエンジン 3 種よりも適訳率はやや高かった。ただし可訳率を合算すると、大半のエンジンがほぼ同等という結果となった。

また、SMT エンジンは RBMT エンジンよりも「訳語のゆれ」が生じる確率がやや高いことも確認された。通常、訳語のゆれは読解の妨げとなり、「粗読」用途や「精読」用途においては悪影響があるが、その反面、「検索」用途においては検索式にヒットする確率が向上するため、有利な面もある（検索ノイズとなる可能性も同時に高まるため、悪影響もある）。

なお、本章冒頭（3.1.1.）で述べたとおり、今回の調査における各エンジンの翻訳条件は一定ではなく、また多くのエンジンは特許翻訳用のチューニングが施されていない。このため、今回の評価結果はあくまで現状の英日機械翻訳の大まかな傾向として把握されるべきものであり、各エンジンの性能の限界値を示すものではない。

3.2. 英日機械翻訳の特有課題・留意点及び対応策

機械翻訳結果の品質評価手法を策定するにあたり、各種機械翻訳における種々の課題や留意点をあらかじめ把握し、これを手法に反映することは不可欠である。このため本調査では、重要技術用語の翻訳精度の評価（3.1）に用いた機械翻訳結果（20 文献）を分析し、機械翻訳における主な課題・留意点を洗い出した。

なお、こうした課題や留意点には、用語の誤訳や係り受けの誤り等、あらゆる言語対において発生し得るものに加え、原語（原文の言語）及び目的言語の特質や、両言語間の特質の違い等により生じる、翻訳言語特有の課題や留意点も存在する。

「仮設手法」では、「1.(3) チェックリスト項目」にて、こうした翻訳言語に応じた課題・留意点もチェック対象としており、中日、韓日、日中、日英のそれぞれについて個別にチェック項目を設定しているが、英日機械翻訳に関しては未設定である。

したがって、本調査にて作成する「特許庁手法」では、英日機械翻訳に関するチェック項目を新たに定義し、ここに追加する必要がある。この目的に資するため、本項にて、上記分析により洗い出した機械翻訳の課題・留意点のうち、英日機械翻訳に特有の課題及び留意点について整理する。

3.2.1. 英日機械翻訳の特有課題・留意点

前項 3.1 にて作成した機械翻訳結果（20 文献×5 エンジン）の分析結果に、過去の知見等を加味して、英日機械翻訳特有の課題・留意点を主要 16 項目に整理した。以下、その内容について示す。

3.2.1.1. 文区切り関連

多くの機械翻訳エンジンは、入力された文の文長が一文として処理可能な限度を超えている場合、文を分割して個別に翻訳し、それらを接続して出力する。この場合、当然ながら分割された部分間に存在した係り受け関係は維持されず、非常に不正確な翻訳となることが多い。

不適切な箇所での文区切りは、このように翻訳結果に大きな悪影響を及ぼすが、英日機械翻訳の場合、さらに以下に示すような課題も存在する。

<EJ-01. 文中のピリオドを適切に処理していない>

文中にてピリオドが記号の一部などで使われた際、これが誤って文区切りとして扱われる等、適切に処理されない場合がある。以下に一例を示す。

原 文	Some signs are hand-held such as the sign taught by U.S. Pat. No. 6,409,358 that uses a flashlight to internally illuminate it.
人手翻訳	標識のあるものは手持ち型であり、たとえば 米国特許第 6, 4 0 9, 3 5 8 号 により教示される標識は内部を照らすために懐中電灯を用いる。
機械翻訳	いくつかの兆候は、 米国特許 により教示印として手持ちされています。内部でそれを照らすために懐中電灯を使用して 番号 6409358 。

上例の文献番号表記「U.S. Pat. No.」にはピリオドが 4 つ含まれており、機械翻訳にて、このうち Pat の後のピリオドが文末と見なされている。

<EJ-02. 「:」や「;」を用いた文を適切に処理していない>

英語特許文献では、「特許請求の範囲」等にて、構成要素を「:」や「;」で区切って一文中に列挙する文が多用される。これに対し、英日機械翻訳では、「:」や「;」を文末と解釈する場合があるが、その際は、前後の係り受けを適切に表現するための配慮が必要である。

原 文	7. Electrically conductive textile yarn (1) according to any one of the preceding claims, wherein said second element (3) consists substantially of fibres selected from the group comprising: aramide fibres, meta-aramide fibres, Lenzing FR(R).
人手翻訳	7. 前記第 2 の要素 (3) が、 以下の 群から選択されるファイバーから実質的になる、先行請求項のいずれか一項に記載の導電性織り糸 (1) : アラミドファイバー、メタアラミドファイバー、L e n z i n g F R (登録商標)。
機械翻訳	7。請求項 1 に記載の導電性の繊維糸 (1)、前記第 2 の要素 (3) が実質的に含む群から選択される繊維 からなる 前記 : アラミド繊維、メタ・アラミド繊維、レンチング FR (R)。

上例の原文は「comprising:」の後に要素が列挙されており、「:」で文を分割する場合はそれを踏まえた対応（人手翻訳における「以下の」等の補足）をする必要があるが、機械翻訳結果は「:」前後の関係性を表現できていない。

3.2.1.2. 単数形・複数形関連

英語には名詞に単数形・複数形の区別があるが、日本語にはない。その結果、英日機械翻訳において、原文にあった数量情報が翻訳結果から欠落してしまう等の問題が生じる場合がある。

<EJ-03. 重要な単数・複数情報が維持されていない>

英文では単数形と複数形の使い分けに大きな意味を持たせることができるが、そのような文において、機械翻訳結果がそのニュアンスを適切に表現できていない場合がある。

原 文	In the embodiment shown the arrangement includes upper and lower pairs of paddles 17, 18 mounted tiltably to the column 11.
人手翻訳	図示された実施態様において、本装置は、角度を変えられることができるように柱 11 に取り付けられた パドル 17、18 の上下 2 対 を含む。
機械翻訳	… パドル 17、18 の上下の対 …

上例の原文は **pairs** と複数形にすることによって、パドル 17、18 の対が上下で 2 対であるということが明白だが、機械翻訳結果ではそのニュアンスが希薄である。

なお、この課題は「単数形と複数形の使い分けに大きな意味を持たせている表現」に限定して問題視すべきものであり、通常の文において単数・複数を区別して訳さないのは日本語の特性によるもので不問とすべきである。

<EJ-04. 単数形・複数形併記表現を適切に処理していない>

英文では、要素が「単数および／または複数」である場合、これを単数形・複数形併記表現（例：an element and/or elements）やカッコ補記表現（例：element(s)）等で表現することがあり、これに適切に対応する必要がある。

原 文	a container according to any preceding claim, wherein a contact member or members is mounted to the container
人手翻訳	前記 1 つ以上の接触部材 は、前記コンテナに取り付けられている、先行請求項のいずれか 1 項に記載のコンテナ。
機械翻訳	前記いずれかの請求項記載のコンテナであり、 接触部材又は部材 がコンテナに装着されているもの

上例の原文 a contact member or members は「1 つまたは複数の接触部材」というニュアンスであるが、機械翻訳結果ではこの表現に対応できておらず、無意味な繰り返し表現となっている。

3.2.1.3. 序数詞関連

冠詞も、英語にあって日本語にない概念のひとつである。冠詞に関連する課題としては次のものが挙げられる。

<EJ-05. 序数詞表現が適切に翻訳されていない>

冠詞に関する特許英文頻出表現として、不定冠詞＋序数詞表現（a first XXX, a second XXX）や、無冠詞＋序数詞表現が多用されることが挙げられる。first、second 等の序数詞は通常定冠詞を伴うものであるため、機械翻訳エンジンによっては、こうした表現に対応できないおそれがある。

原 文	Heat shield 7 to this end has a first channel 71 and a second annular channel 72 receiving, with interposition of a sealing strip, the walls of upper part 9.
人手翻訳	熱シールド7は、この目的のために、第1チャンネル71と、シーリングストリップを介在させて上側部分9の壁を受け入れる第2の環状チャンネル72とを有する。
機械翻訳	第1は、このために熱シールド7で71と2分の1の環状流路72受信にチャンネルを開設する、シールストリップの挿入物で、上部9の土手

上例では、「第2の」と訳すべき a second が「2分の1の」と誤訳されている。second は「秒」と誤訳されることも多い。ここでは a first の係り受けも誤っている。

3.2.1.4. 語順に関する課題

英文の語順に起因する英日機械翻訳の課題を列举する。

<EJ-06. 慣用表現への語の挿入を適切に処理していない>

英文においては、複数の語から成る慣用表現の中間への語の挿入が一般的に行われるが、英日機械翻訳では、こうした語の挿入が原因で慣用表現をそれと認識できずに誤訳してしまう場合がある。

原 文	For example, sacrificial anode layers may be placed in electrical contact with metal surfaces to be protected from corrosion.
人手翻訳	例えば、犠牲アノード層は、腐食から保護すべき金属表面と電氣的に接触して配置することができる。
機械翻訳	例えば、犠牲陽極層は腐食から保護される金属表面を備えた電気接点に置かれるかもしれません。

in contact with は「～に接して」という慣用表現だが、上例ではここに electrical が挿入されている。機械翻訳結果は、これを慣用表現への語の挿入と認識できず、個々の語に分解して翻訳したため、誤訳となっている。

<EJ-07. 倒置文を適切に翻訳していない>

英語特許文献では、主語が長大になる場合に文意を理解しやすくする等の目的で、倒置文が多用される。

原 文	Spanning the width at the base of container 1 are three slide bars 9,10, 61 each having low friction hard upper surfaces.
人手翻訳	コンテナ 1 の基部において、低摩擦硬質上面をそれぞれ有する 3 つのスライドバー 9、10、61 が幅の方向に横切っている。
機械翻訳	コンテナ 1 の基礎で幅を測ることは、低摩擦を各々困難にしておく 3 本のスライドバー 9、10、61 である、上部の表面。

原文は **are** の前後が入れ替わった倒置文であるが、機械翻訳結果ではこれを正しく解釈できておらず、冒頭の Spanning～を主語と誤認している。

<EJ-08. 名詞と代名詞／省略表現の関係を適切に表現できていない>

同一文中にて既出の名詞が再度出現する場合、代名詞や省略表現（例：machine translation system であれば the system 等）が用いられることが多いが、文の構造によっては、機械翻訳結果にて両者の順序が入れ替わり、関係性がわかりにくくなる場合がある。

原 文	Paddles 17, 18 tiltably supported by the restraint 11 may engage the items to secure them .
人手翻訳	角度を変えられることができるように固定具 11 によって支持されたパドル 17、18 が 品物 と係合して 品物 を確保することができる。
機械翻訳	拘束具 11 により傾斜して支持されるパドル 17、18 は、 それら を固定するために 物品 を拘束してもよい。

上例原文末尾の **them** は **items** を指すが、機械翻訳結果では **them** のほうが先に出現してしまうため、語順的にはパドル 17,18 を参照しているように解釈されてしまう。この課題は、英日の言語特性の違いに起因する課題であり、上記翻訳結果も逐語訳としては誤りとは言えないが、理想的にはこうした特性の差異を踏まえた訳出が望まれる。

3.2.1.5. 多品詞・多義語関連

多品詞・多義語を原因とする翻訳ミスはあらゆる言語対において発生し得るものであり英日特有の課題とはいえないが、英語は多品詞や多義語が特に多いため、一層の留意が必要である。

ここでは、多品詞や多義語による翻訳ミスの一種であるが、特に英日機械翻訳に顕著と思われるものについて述べる。

<EJ-09. 慣用表現とみなすべき場合／そうでない場合を適切に翻訳できていない>

同一の表現が慣用表現として使われている場合とそうでない場合とがあるため、文意文脈に応じて正しく解釈する必要がある。

原 文	This facilitates loading and unloading the container and also makes the container adaptable to carry differing quantities of windscreens or windscreens of differing size and curvature configuration.
人手翻訳	これはコンテナの荷積みおよび荷降ろしを容易にし、コンテナをさまざまな 数量の フロントガラス、またはさまざまなサイズおよび曲り方の構成のフロントガラスを運ぶのに適したものとする。
機械翻訳	…異なる 多量の フロントガラスを運ぶのに適応…

上例は、文意を考えると **quantities of** は「(さまざまな) 数量の (フロントガラス)」と訳すのが正しいが、機械翻訳結果では「**quantities of** : 多量の」という慣用表現に解釈されてしまっている。一方、下例のように、慣用表現として訳すべき語がそう訳されない場合もあり得る。

原 文	[0002] A number of devices have been proposed for suspending washed laundry for drying.
人手翻訳	【 0 0 0 2 】洗濯された洗濯物を乾燥のためにかけるための 複数の 装置が提案されている。
機械翻訳	【 0 0 0 2 】装置の 数 は、乾燥のための洗浄洗濯物を懸架する提案されている。

A number of は「複数の、多数の」という意味の形容詞となる慣用表現だが、機械翻訳ではそう解釈せず、**number** を単独の名詞と解釈してしまっている。

<EJ-10. 前置詞に対し適切な訳語を選択できていない>

of や for などの前置詞の適訳は、述語や目的語の属性等により変化する。例えば for であれば、述語や目的語に応じて「～のために」「～用の」「～という理由で」「～に向かって」など多数の訳語に訳し分ける必要がある。

原 文	The cylindrically curved controllable platform 128 can be rotated about its axis in either a forward or backward direction to provide material deposition in either direction.
人手翻訳	シリンダ状に湾曲した可変台 128 は、その軸の 周囲 を前後どちらの方向にも回転して両方向に材料を被覆することができる。
機械翻訳	一方の方角に資料被覆を提供するために、円筒状に曲がった制御可能なプラットフォーム 128 は、前方あるいは逆方向のいずれかの中のその軸芯 に関して 回転することができる。

上例原文の前置詞 about は、述語 rotate を受け、ここでは「～の周囲を」等と訳すべきである。なお、この翻訳ミスは、rotate about を慣用表現とみなせば前出の EJ-09 に該当するものともいえる。

3.2.1.6. 造語関連

英語文は、語間がスペースで区切られているため、語の認定が容易である反面、英日機械翻訳においては、語の前後に文字を付け足したり、複数の語を連結したりすると、既存の語と認識されず、未知語扱いされたり、適切に翻訳されない場合がある。以下、これに該当する英日機械翻訳の課題・留意点を挙げる。

<EJ-11. 接頭辞、接尾辞を伴う語を適切に処理していない>

英語特許文献には、任意の語の前後に接続して意味を追加する接頭辞・接尾辞が多用される。主な例としては de-XXX、non-XXX、XXX-free、XXXly などが挙げられるが、こうした接頭辞・接尾辞の使用により、語が正しく翻訳されなくなる場合がある。

原 文	The invention relates to a testing method for designing a semiautogenous or an autogenous grinding circuit with at least one ball mill for grinding ore.
人手翻訳	本発明は、少なくとも 1 つのボールミルを有する 半自生 粉砕回路または自生粉砕回路を設計するための試験方法に関する。
機械翻訳	本発明は研磨する鉱石のために少なくとも 1 個のボールミルで semiautogenous または自生粉砕回路を設計するための試験法に関する。

上例では、原文中の autogenous は「自生」と適切に翻訳されているが、接頭辞 semi を伴う semiautogenous は未知語となってしまっている。

<EJ-12. スラッシュを用いた並列表現を適切に処理していない>

英語特許文献では、スラッシュ (/) を用いて名詞や動詞、前置詞等を並列させる表現も多用される。

原 文	11. A perimeter barrier according to any preceding claim, wherein one or more additional wire ropes is/are provided, unwoven or woven relative to the intermediate post.
人手翻訳	1 1. 前記間柱に対して編まれていないかまたは編まれた、1 つまたはそれ以上の追加のワイヤロープが提供される、先行する請求項のいずれかに記載の周辺防壁。
機械翻訳	1 1.いかなる以前の請求項に記載の周辺部バリアも、そこにおいて、中間ポストと関連して設けられているか、ほどかれるかまたは、織られる一つ以上の付加的なワイヤロープ i s / a r e 。

原文の動詞 is、are は基本語だが、両者がスラッシュで連結されたことによって機械翻訳では一語とみなされ、未知語扱いされてしまっている。

3.2.1.7. 英語特許文献の頻出表現に関する課題・留意点

特許文献にて多用される表現や文型、文フォーマットには、一般の文ではあまり使用されないものも少なくない。こうした傾向は英語に限らないため英日特有の課題とはいえないが、ここでは、英語特許文献にて多用される表現や文型、文フォーマットのうち、特に英日機械翻訳にて適切に処理されない危険性が高いものについて示すことにする。

<EJ-13. therein、wherein、whereby 等を適切に処理していない>

「発明の名称」部・「特許請求の範囲」部をはじめ各所にて、therein, therefor, wherein といった自然文ではあまり用いられない副詞が多用される。こうした副詞を適切に表現する必要がある。

原 文	HYDROGENATION CATALYSTS AND PROCESSES THEREWITH
人手翻訳	水素添加触媒および これを用いた 工程
機械翻訳	水素化触媒とプロセス それを用いた

特に **wherein** は、英語特許文献の「特許請求の範囲」部分における定型文に含まれ多用される重要な用語である。

原 文	46. The method of any of claim 43-45, wherein the sequence patterns are nucleic acid motifs, amino acid motifs or protein domains.
人手翻訳	【請求項 4 6】 配列パターンが核酸モチーフ、アミノ酸モチーフ又はタンパク質ドメインである、請求項 43-45 のいずれかに記載の方法。
機械翻訳	46. 請求項 4 3-4 5 (シーケンス・パターンが核酸のモチーフ、アミノ酸のモチーフまたはタンパク質領域である) のいずれかの方法。

上例では、**wherein** 以下をカッコ表記にしており、これ自体は意味の把握を阻害するものではないが、カッコ表記が「請求項 4 3-4 5」の説明としてこれの直後に挿入されている点が不適切である。

<EJ-14. 代名詞 **same**、**the same** を適切に処理していない（【発明の名称】）>

英語特許文献では、「発明の名称」部において **same** や **the same** の代名詞用法が多用される。英日機械翻訳では、この用法に適切に対応できていない場合がある。

原 文	Combination pouch and release sheet and method for producing same
人手翻訳	組合せ式パウチ、剥離シート、及び その 製造方法
機械翻訳	コンビネーションポーチ、分離型シート、および 同じ を生産する方法

上例では **same** は代名詞として使用されているが、機械翻訳結果では「同じ」という訳語から判断して、これを形容詞と解釈している可能性が高い。なお、**same** を代名詞として直訳すると「同じもの」「同じこと」となるが、特許文献の機械翻訳では、理想的には人手翻訳のように「その」等の自然な表現とすべきである。

<EJ-15. 小文字を含む大文字表記を適切に処理していない（【発明の名称】）>

英語特許文献の「発明の名称」は全文字大文字で表記されることが多いが、まれに、全文字大文字表記でありながら、元素記号など一部のみ小文字が用いられるケースがある。このような特殊なケースが、全文字大文字表記のバリエーションと判定されず、適切に処理されない懸念がある。

原 文	BACTERIUM COMPRISING A REGULATED rfaH NUCLEIC ACID
人手翻訳	調整された rfaH スクレイン酸からなるバクテリア
機械翻訳	BACTERIUM COMPRISING A REGULATED rfaH スクレイン ACID

上記例では、**rfaH** が小文字を含んでいるため、原文が全文字大文字表記とみなされず、大文字表記の大半が翻訳不要語として扱われた。原文の **rfaH** を大文字にして再度機械翻訳すると「規制された **RFAH** 核酸を含む細菌」と訳されることから、**rfaH** が小文字を含むことによる翻訳ミスであると判断される。

<EJ-16. 引用クレームに関する表現を適切に翻訳していない（【特許請求の範囲】）>

「特許請求の範囲」において、前出の請求項を引用するいくつかの定型表現がある。特許文献の英日機械翻訳では、こうした表現に適切に対応する必要がある。

原 文	10. a container according to any preceding claim , wherein a lower portion of the container includes frame openings for receiving tines of lifting apparatus.
人手翻訳	【請求項 10】 前記コンテナの下側部分は昇降用装置の先端部を受け入れるためのフレーム開口を備える、 先行請求項のいずれか 1 項に記載の コンテナ。
機械翻訳	10. どんな前のクレームに記載の コンテナ; そこでは、コンテナの下位部が浮き器具のティンを受けるためのフレーム開始を含んでいる。

3.2.2. 英日機械翻訳特有課題への対処策

3.2.1.にて挙げた英日機械翻訳特有の課題と留意点は、その大半が機械翻訳エンジンのロジックに直接関係するものであり、特に RBMT 方式に関しては、エンジン内部の調整が必要となるものが多い。ただし、ユーザレベルでの対応としても、造語関連の頻出語をユーザ辞書に登録する、「発明の名称」や「特許請求の範囲」における頻出の構造をパターン登録する等の対処は、多くのエンジンにおいて可能である。

一方、SMT 方式に関しては、学習コーパスをユーザ側で選定できる場合は、それぞれの課題に該当する文を多数学習させることで一定の効果が得られる可能性がある。ただし SMT 方式の性質として、対処の結果どの程度ユーザの思惑通りの出力が得られるや、どのような副作用が生じるかが RBMT 方式に比べて予測しにくい傾向にあるため、その点への留意が必要である。

第 4 章

特許庁機械翻訳品質評価手法の作成

第4章 特許庁機械翻訳品質評価手法の作成

本調査では、「特許庁における機械翻訳品質評価に関する基本的な考え方」(添付資料 1-1)、並びに先に実施した文献調査(第2章)、英日機械翻訳の現状分析(第3章)の調査結果を踏まえ、「仮設機械翻訳品質評価手法(仮設手法)」(添付資料 1-2)の内容を再検討して、「特許庁機械翻訳品質評価手法(特許庁手法)」の素案となる「新設機械翻訳品質評価手法」を作成した。

その後、機械翻訳に関する専門家、翻訳品質評価に関する専門家、特許文献の翻訳に関する専門家等で構成される委員会(第1回委員会¹⁾)にてその妥当性を諮り、そこで得られた指摘及び知見に基づき内容を修正、これを「特許庁手法」として定めた。

以下、本調査にて作成した「特許庁手法」の各項目の内容について、作成時に考慮した検討事項や委員会における指摘事項等を含めて説明する。なお、「特許庁手法」全体の内容は添付資料 1-3 に示す。

4.1. 特許庁機械翻訳品質評価手法の前提条件

「特許庁手法」の作成にあたり、まずはその前提とすべき条件について、「特許庁における機械翻訳品質評価に関する基本的な考え方(以下、「基本的な考え方」という)」の定義、及び「仮設手法」の記載内容等も踏まえ、以下のとおり整理した。

4.1.1. 機械翻訳の対象

機械翻訳の対象は特許文献である。したがって、「特許庁手法」の評価対象は、特許文献の機械翻訳結果となる。ただしその評価単位は、特許文献から抽出され「テストセット²⁾」に含められた単独の文単位となる。

4.1.2. 機械翻訳結果の用途

「特許庁手法」にて評価する機械翻訳結果の用途は、以下のとおりである。

- ① 検索用途：先行技術調査におけるテキスト検索用のデータとしての利用。
- ② 粗読用途：先行技術のあたりをつけるための粗読用データとしての利用。
- ③ 精読用途：先行技術の内容を正確に把握するための精読用データとしての利用。

機械翻訳結果は、特許庁内で使われている検索システムのコンテンツとして、上記の各

¹ 第1回委員会の主な提言・指摘事項は、添付資料 1-4 を参照。

² 「特許庁手法」では、特許文献から選定した文をあらかじめ「テストセット」として保有し、評価にあたってはここから評価対象文を抽出してこれを機械翻訳する。詳しくは 4.4. を参照。

用途に供され、ユーザ（審査官）は、基本的に日本語の文献と同様の手法によって本コンテンツを検索することを想定する。

上記の各用途は一連の業務として行われる場合が多いが、「特許庁手法」においては、「検索」とは「データベースから機械翻訳結果を検索する」ところまでを指し、一方、「粗読・精読」は「（すでに手元にある）機械翻訳結果を粗読／精読して内容を理解する」部分のみを指すものと区分する。

この区分に基づく、「検索」用途においては、その文献がテキスト検索でヒットするかどうかのみが適否評価の観点となる。このため技術用語の翻訳精度が最も重要となる。ただし、特許実務では技術用語の出現順序や付随する助詞等も考慮されるため、文としての内容の正確さについても評価に加える必要がある。

一方、「粗読」及び「精読」用途の場合、技術用語の翻訳精度や文の正確さは同じく重要であるが、最低限、その文献が調査対象とすべきものかどうか判断できればある程度の有用性をもつため、検索用途と比較すると、その適否評価はより段階的かつ多角的な観点で行うべきものである。例えば、文の流暢さは「検索」用途にはほぼ無関係であるが、「粗読」及び「精読」用途においては、補助的な観点として考慮が必要となる。

4.1.3. 品質評価の目的・利用方法

「特許庁手法」を用いて実施する品質評価の目的・利用方法としては、以下の 3 つが想定される。

① 相対評価

複数の機械翻訳システム間の翻訳精度比較や、特定のシステムの翻訳品質の定点観測（同一システムの二つの時点の翻訳精度比較）等のための評価。

② 絶対評価

評価対象とした機械翻訳結果の特定の用途への有用性の判断等のための評価。具体的な用途として、検索、粗読のほか、実体審査での精読、権利侵害の存否等も想定される。

③ フィードバック利用

機械翻訳システムの具体的な弱点の把握と品質向上等のためのフィードバックのための評価。

「特許庁手法」は、このうち最も喫緊の目的である「相対評価」に主眼を置きつつも、「絶対評価」及び「フィードバック利用」にも充分に利用可能なものとする。

なお、「相対評価」と「絶対評価」は1回の品質評価の結果が双方に利用される可能性を想定するが、「フィードバック利用」のための評価は、これらとは切り離して別途行う前提とする。このため、「フィードバック利用」のための品質評価の手法は、「相対評価」「絶対評価」と共通である必要はない。

4.1.4. 評価者

「特許庁手法」の評価者は、特許文献とその利用に対する知識があり、かつ日本語³に充分精通してあることを前提とする。ただし、上記条件に該当し、かつ高度な機械翻訳知識や文法知識、原文読解力を兼ね備える人材は確保が難しいため、「特許庁手法」では、原則として、これらの知識・能力を必須とする評価項目は設けないものとする。

ただし、フィードバック利用のための評価に関しては、原文と機械翻訳結果とを照合して翻訳ミスの内容と原因を特定する作業が必須となるものと想定されるため、この評価用に設ける評価観点に限り、評価者の要件として原文読解力を前提とすることも許容する。

4.1.5. 評価作業のリソース制限

品質評価作業に充当可能なリソース、具体的には評価者の人数、作業時間、コスト等には一定の制限が生じる可能性が高い。このため、「特許庁手法」は、品質評価に要する作業負担や費用負担が過大とならぬよう配慮し、かつ十分なリソースが確保困難な場合の対処方法等についても考慮する。

³ 日本語が目的言語とならない場合は、目的言語に応じた評価者を起用する必要がある。

4.2. 特許庁機械翻訳品質評価手法の構成

「特許庁手法」のベースとなる「仮設手法」の構成は下表 4-1 のとおりである。

表 4-1. 「仮設手法」の構成

> 事前準備
> 人手による品質評価の実施
1. 評価の観点及び基準
(1) 内容の伝達レベルの評価
(2) 技術用語に関する誤訳の個数
(3) チェックリストに基づく評価
2. 相対評価方法
3. 絶対評価方法

「特許庁手法」では、上記「仮設手法」の構成に準拠しつつも、これを表 4-2 のとおり改めた。

表 4-2. 「特許庁手法」の構成

> 事前準備
> 人手による品質評価の実施
1. 評価の観点及び基準
(1) 内容の伝達レベルの評価
(2) 重要技術用語の翻訳精度の評価
(3) 流暢さの評価
(4) チェックリストに基づく評価
・ 必須項目
・ 任意項目
・ 翻訳言語別任意項目
2. 相対評価方法
3. 絶対評価方法
4. 主要項目と補助項目

※赤字箇所は「仮設手法」から変更された箇所を示す。

上表中、「仮設手法」の構成を変更・追加した項目は赤字で示している。なお、構成上は変更のない項目であっても、各項目中の定義内容は全て見直し、修正を加えている。

次項より、「特許庁手法」各項目について、検討の経緯と、その結果確定した定義内容を、「仮設手法」における課題、文献調査等にて得た知見、委員会での指摘等に照らして詳述する。

なお、「特許庁手法」の構成は、上表 4-2 のとおり、最初に「事前準備」部分、続いて「人手による品質評価の実施」部分、という記載順序になっているが、「事前準備」部分で定義される内容は、評価に用いるテストセットの仕様等、すべて具体的な評価観点や評価基準を定めたうえで検討すべきものである。このため、「特許庁手法」の検討は、評価観点や評価基準を定義している「人手による品質評価の実施」部分から着手し、その後、その内容を踏まえて「事前準備」部分を見直した。

本報告書もこの検討順序に合わせ、まずは次項 4.3.にて「人手による品質評価の実施」部分を、続く 4.4.にて「事前準備」部分を説明する。

4.3. 特許庁機械翻訳品質評価手法の内容（1）：「人手による品質評価の実施」部分

「人手による品質評価の実施」部分は、「特許庁手法」として実施する品質評価の観点及び基準を定義し、かつ相対評価方法、絶対評価方法等、品質評価結果の利用方法について定める中核部分である。

「特許庁手法」の「人手による品質評価の実施」部分の検討は、これを構成する「1. 評点の観点及び基準」「2. 相対評価方法」「3. 絶対評価方法」のそれぞれについて個別に行なった。さらに、これら「仮設手法」から引き継いだ項目に加え、新たに「4. 主要項目と補助項目」を追加した。

また、「1. 評点の観点及び基準」については、「仮設手法」にて本項目を構成していた「内容の伝達レベルの評価」「重要技術用語の翻訳精度の評価」「チェックリストに基づく評価」の3つの評価観点に加え、新規の観点として「流暢さの評価」を追加した。

以下、これら「人手による品質評価の実施」部分の各項目について、その検討内容を詳しく説明する。

4.3.1. 評価の観点及び基準：(1) 内容の伝達レベルの評価

「仮設手法」の「内容の伝達レベルの評価」部分の記載内容は以下のとおりである。

表 4-3. 「仮設手法」「(1)内容の伝達レベルの評価」部分

<評価基準に基づき、1－5点の評価を行う> 5：すべての重要情報は正確に伝達されている。 4：ほとんどの重要情報は正確に伝達されている。 3：翻訳は、完全に理解できるというわけではないが、いくつかの重要情報は正確に伝達されており、正確でない情報についても大まかな内容理解が可能である。文法的、文体的に不自然な部分もあるが、内容は十分に反映されている。 2：いくつかの重要情報は正確に伝達されており、正確でない情報についても、他の文章や図面の参酌、十分な検討の時間等が与えられれば、大まかな内容を理解できる可能性がある。 1：翻訳は全く理解できない。正確に伝達されている情報はほとんどない。 <評価対象件数について、複数人（3人以上）による重複評価を行う> <評価は評価対象となる各案件の属する各技術分野に精通した評価者が行う> <評価に先立って、サンプルデータ及びサンプル評価の意識合わせを行う> <複数人による評価結果の平均値を算出し、さらに評価対象件数の平均値を算出する> <平均値を、内容伝達レベルの評価値とする>

「仮設手法」の評価基準は、表 4-3 に示したとおり、原文に含まれる重要情報が機械翻訳結果においてどの程度正確に理解可能であるかを 5 段階で評価する形式である。この「機械翻訳結果が原文の内容を正確に反映しているか否か」、すなわち「内容の正確さ」という評価観点は、過去の多くの人手評価手法においても Adequacy 等の形で多用されている主要な評価観点のひとつであり、また第 2 章 2.1.5.1. で述べたとおり、特許文献機械翻訳結果の実務への有用性を評価するうえでは最も重要な評価観点といえる。

したがって、「特許庁手法」においても、「内容の伝達レベルの評価」は引き続きメインの評価観点として採用した。ただし、採用にあたっては、以下の各点について再検討のうえ、定義内容への加筆修正を行なった。

4.3.1.1. 流暢さの観点の排除

「流暢さ (=Fluency)」とは、機械翻訳結果の文章としての読みやすさ、理解し易さを評価する観点である。通常、機械翻訳結果を独立した文として扱うため、原文の意味の伝達度は考慮されない。

「流暢さ」の評価は、通常は「内容の正確さ (Adequacy)」の評価とは別途に行われることが多く、このため「内容の正確さ」の評価の際に「流暢さ」が考慮されることは少ないが、「内容の正確さ」の観点に「流暢さ」の要素を加味し、総合評価的に運用される場合もある。

「特許庁手法」の「内容の伝達レベルの評価」に「流暢さ」の要素を加味すべきか否かについては、機械翻訳結果の主用途における「流暢さ」の重要度を考慮して決定する必要がある。4.1.2. に示したとおり、「特許庁手法」の想定する機械翻訳結果の主要な用途には「検索」用途と「粗読／精読」用途とがあり、前者においては「流暢さ」はほぼ無関係であるが、後者においては補助的な観点として考慮が必要な要素となる。したがって、「特許庁手法」にてこれらの用途への有用性を評価する際には、それぞれで「流暢さ」の扱いを変えなければならない。具体的には、「検索」用途に関する評価においては「流暢さ」は排除すべきであり、一方「粗読」用途や「精読」用途に関する評価においては、これを補助的な観点として考慮する必要がある。

前記した「内容の正確さ」に「流暢さ」の要素を加味する総合評価方式は、評価結果をそれぞれの要素に分解することができないため、上記運用に対応できず、「特許庁手法」には適していない。したがって、「特許庁手法」では、「内容の伝達レベルの評価」に「流暢さ」の要素は加味するべきではないと判断した。

「仮設手法」における評価基準も、評点 5 や評点 4 の定義を見る限り、「流暢さ」の要素は考慮していないと考えられるが、その一方で、評点 3 では「文体的な自然さ」も評価観

点に含めており、やや曖昧な部分もある。「特許庁手法」では、評価基準からこうした曖昧な記述を除去し（4.3.1.2.にて後述）、さらに「流暢さ」を評価観点に含めないよう、以下のとおり注記して、方針を明確に示した。

※機械翻訳結果の「流暢さ」については「(3)流暢さの評価」で評価するため、ここでは考慮しない。

なお、「流暢さ」は、「内容の伝達レベルの評価」からは排除するが、前述のとおり「粗読」用途や「精読」用途への有用性の評価においては必要な観点である。「仮設手法」には「流暢さ」を評価する観点が存在しないため、「特許庁手法」にてこれを新設する必要がある（4.3.3.にて後述）。

4.3.1.2. 各評点の定義の見直し

「仮設手法」の 5 段階の評価基準は、先に述べたとおり、評点 3 の定義中に「流暢さ」の観点と見なされる記述があり、「特許庁手法」ではこの部分を排除する必要があるが、それに加え、評点 2 のみが「他の文章や図面の参酌、十分な検討の時間」等の補助情報の使用を仮定しており、他の評点と前提条件が異なるため、評価時に混乱を招くおそれがある。仮に補助情報の利用を許容するならば、全ての評点において許容すべきである。ただし、その補助情報についても、評価者によって想定する内容に差が生じないように、具体的な内容や品質を定義しなければならず、徒に評価基準を複雑にする懸念がある。

また、評点 3 や評点 2 のように定義に具体的な条件を数多く挙げると、それに完璧に合致した場合の判断は容易になる反面、複数の評点に部分的に合致するというケースが多発し、評価者が評点を決めにくくなるという問題も生じてくる。

このため「特許庁手法」では、各評点の定義を「仮設手法」における評点 5 や 4 のようなシンプルなフォーマットに揃え、評点 3 や 2 で挙げられている諸条件は廃止することとした。以下、その定義内容を示す。

- 5：すべての重要情報が正確に伝達されている。
- 4：ほとんどの重要情報は正確に伝達されている。
- 3：半分以上の重要情報は正確に伝達されている。
- 2：いくつかの重要情報は正確に伝達されている。
- 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。

4.3.1.3. パーセンテージ表示の併記

「特許庁手法」における各評点の基準は、4.3.1.2.に示したとおり、具体的な条件を廃止したシンプルな定義とした。ただし、これらの定義に含まれる「ほとんど」や「いくらか」という表現は、評価者によって捉え方が異なる場合があり、評価者間で評価の水準、すなわち厳しき／甘さにずれが生じる可能性がある。

こうした評価水準のずれへの対応は、評価サンプルの提示や評価者間の意識合わせ等、運用面での対処が必須であるが、評価基準においても、評価者に大体の目安を提示する方策として、第 2 章で調査したいくつかの品質評価基準で導入されていた「パーセンテージ表示の併記」が有効であると判断し、これを導入した。

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%)

4.3.1.4. 重要情報の定義

「仮設手法」では、評価基準に「重要情報」という表現が用いられているが、これが何を指すかが定義されておらず、評価基準のずれの一因となり得る。「特許庁手法」では、評価基準にこの「重要情報」の定義を追加し、その内容を明確化した。

※「重要情報」とは、原文に含まれる全ての実質情報（技術的要素とその相互関係）を指す。各重要情報の重要度と訳語の精度から、総合的な評価を行う。なお各評点に付記したパーセンテージは、重要情報が正確に伝達されている度合いの大体の目安である。

なお、「内容の伝達レベルの評価」は文全体を対象とした主観的な評価を意図したものであるが、上記定義及びパーセンテージ補記により、評価者に「文中の技術的要素とその相互関係を数え上げ、このうち正確なものの比率を算出してパーセンテージに当てはめる」評価との誤解を招く懸念がある。これを防止するため、本評価の主旨を評価基準の冒頭に以下のとおり明記した。

(1) 内容の伝達レベルの評価

機械翻訳結果が、原文の内容をどの程度正確に伝達しているかを、人手翻訳の内容に照らして以下の 5 段階の評価基準で主観的に評価する。

4.3.1.5. 補記部分

「仮設手法」には「内容の伝達レベルの評価」の実施要件を定義する 5 つの補記がなされている。

1 つめの補記では、「内容の伝達レベルの評価」を、3 名以上の評価者による重複評価に行うことが定められている。他の評価観点にはこの定めはなく、「内容の伝達レベルの評価」のみの要件となる。

この定義は、「内容の伝達レベルの評価」が機械翻訳結果の「内容の正確さ」を測るものであり、特許文献の機械翻訳結果の有用性を評価するうえで最も重要な観点の評価であること、そしてその反面、主観的な評価であり、評価者間で評価基準のずれが生じやすい性質をもつことを考えると適切であり、複数の評価結果の平均値を得ることで、より正確かつ客観的な評価結果が得られると考えられる。このため「特許庁手法」においても必須の要件とする。

2 つめの補記は、それぞれの評価対象文を、その文の属する技術分野に精通した評価者に担当させることを要件とするものである。「内容の伝達レベルの評価」は、文に含まれている技術内容が正確に読み取れるか否かを評価するものであるため、当該技術分野の知識は必須であり、こちらも妥当な要件である。

ただし、この要件と、1 文に対し 3 名以上で重複評価するという 1 つめの要件とを考え合わせると、全技術分野を対象とした評価を行う場合には、相当数の評価者を用意する必要が生じ、その確保が困難な場合も想定される。このあたりは、都度のリソース制限に応じて柔軟に運用できるような定義が別途必要であろう。この点は 4.3.10.2. にて検討する。

なお、第 1 回委員会にて、複数のシステムを相対評価する場合、公平な評価とするためには、同一の文については必ず全てのシステムの機械翻訳結果を同一の評価者が同時に評価する必要があることが指摘され、「特許庁手法」にこれを明示することが求められた。この指摘を受け、2 つ目の補記の末尾に以下の文言を加えた。

1 つの機械翻訳対象文に対し複数の機械翻訳結果を評価する場合は、その全てを同一の評価者が同時に評価する。

この指摘は本観点に限らず、全ての評価観点に該当するものであるため、「特許庁手法」では全ての評価観点にこの補記を追加した。

3 つめの補記は、評価者間の評価の水準（厳しさ／甘さ）のずれを抑えるために、評価に先立ってサンプルデータ及びサンプル評価の意識合わせを行うことを定めたものである。

「内容の伝達レベルの評価」の各評価基準、例えば評点4の示す「ほとんどの重要情報」の「ほとんど」がどの程度の範囲を指すかを明確に定義することは困難であり、最終的には評価者の主観に委ねざるを得ない。「特許庁手法」では、評価者間の水準のずれを軽減するため、各評点にパーセンテージを併記して判断の目安としたが、この補記に定められているように、各評価者間でサンプルを用いて意識合わせを行うことも、評価者間の水準を合わせるのに非常に有効と考える。

なお、評価水準の意識合わせは、その都度の評価者間のみでなされれば充分とは限らず、例えば1つの翻訳システムに対して経時的変化（例えばエンジンのバージョンアップによる精度向上）の確認を行うような場合は、前回の評価水準との整合が必要となる。このため、初回の意識合わせにて実施したサンプル評価の結果を保存し、次回以降のリファレンスに利用することが推奨される。これを踏まえ、「特許庁手法」では3つめの補記に対して以下のとおり加筆した。

＜評価に先立って、サンプルデータ及びサンプル評価の意識合わせを行う。**なおサンプル評価結果は、意識合わせの内容を反映させたうえで、次回以降のリファレンスとして保存・活用する**>※

※この補記については9.1.1.にて更に加筆修正している。

4つめ、5つめの補記は、1つめの補記で定義した「複数名での重複評価」の結果の扱いに関するものであるが、評価結果の取扱いについては、後述の「相対評価方法」（4.3.8.参照）にて他の評価観点と一括して定義することとしたため、本観点からは削除した。

4.3.1.6.「特許庁手法」の確定内容

上記の再検討結果を踏まえ、「特許庁手法」における「内容の伝達レベルの評価」の内容を、以下のとおり確定した。

表 4-4.「特許庁手法」「(1)内容の伝達レベルの評価」

(1) 内容の伝達レベルの評価

機械翻訳結果が、原文の内容をどの程度正確に伝達しているかを、人手翻訳の内容に照らして以下の 5 段階の評価基準で主観的に評価する。

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%)

※「重要情報」とは、原文に含まれる全ての実質情報（技術的要素とその相互関係）を指す。各重要情報の重要度と訳語の精度から、総合的な評価を行う。なお各評点に付記したパーセンテージは、重要情報が正確に伝達されている度合いの大体の目安である。

※機械翻訳結果の「流暢さ」については「(3)流暢さの評価」で評価するため、ここでは考慮しない。

<評価対象件数について、複数人（3人以上）による重複評価を行う>

<評価は評価対象となる各案件の属する各技術分野に精通した評価者が行う。1つの機械翻訳対象文に対し複数の機械翻訳結果を評価する場合は、その全てを同一の評価者が同時に評価する>

<評価に先立って、サンプルデータ及びサンプル評価の意識合わせを行う。なおサンプル評価結果は、意識合わせの内容を反映させたうえで、次回以降のリファレンスとして保存・活用する>

4.3.2. 評価の観点及び基準：(2) 重要技術用語の翻訳精度の評価

本評価観点は、「仮設手法」の「(2) 技術用語に関する誤訳の個数」に相当する。「仮設手法」の定義は以下のとおりである。

表 4-5. 「仮設手法」の「(2)技術用語に関する誤訳の個数」

・ 誤訳の語数のカウント

対象文献の技術分野を鑑み、特許文献において重要と想定される技術用語の訳が、人手翻訳に対して技術的に同義な範囲を外れているもの（未知語、訳漏れを含む）の数をカウントする。

＜技術用語のチェックは、評価対象となる各案件の属する各技術分野に精通した評価者が行う＞

なお、本評価の対象となる「特許文献において重要と想定される技術用語（以下「重要技術用語」という。）」は、全ての翻訳対象文についてあらかじめ特定されている前提であり（4.4.1.4.）、評価者自身がこれを選定する必要はない。

4.3.2.1. 適訳／誤訳基準の見直し

「仮設手法」における重要技術用語の評価基準は、上に示したとおり、「人手翻訳に対して技術的に同義な範囲を外れている」訳語を一律で誤訳と見なし、その数をカウントするというものである。すなわち、同義であれば「適訳」、同義でなければ「誤訳」の 2 段階評価である。

この 2 段階評価においては、機械翻訳結果に数多く含まれる「一般的に使われる訳語でなく同義の範囲外だが、正しい意味は分かる」という訳語（いわば「可訳語」）も「誤訳」と判定され、完全な誤訳語や未知語・訳漏れと同等の扱いとなる。確かに、可訳語は「検索」用途においては有用性に乏しく、誤訳語と同等の評価が妥当であるが、その一方で、原文の内容を把握するための「粗読」用途や「精読」用途においては十分な有用性があり、むしろ適訳語に近い評価が必要である。

「仮設手法」の 2 段階評価は、完全な誤訳語と可訳語とを区別していないため、こうした用途別の調整ができない。本調査の「英日機械翻訳の現状分析（第 3 章）」では、この点を問題視し、技術用語の誤訳率の確認の際、「仮設手法」の評価基準を採用せず、新たな評価基準を用いた（3.1.4.）。この新たな評価基準では、「適訳語（A）」と「誤訳語（C）」の間に「可訳語（B）」が追加され、用途に応じて可訳語の扱いが変更可能となっている。また、未知語・訳漏れを「不訳語（D）」とし、こちらも「誤訳語（C）」と区別できるようにしている。

この新評価基準は、もともと「特許庁手法」への採用を想定して考案したものである。「英

日機械翻訳の現状分析」の実施を通じ、新評価基準が良好に機能することが確認できたため、予定どおり「特許庁手法」に採用することとした。

- A（適訳）：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）：技術用語として一般的に用いられる訳語ではないが、意味はおおむね正しい。
C（誤訳）：誤訳である。
D（不訳）：未知語、訳漏れである。

4.3.2.2. 誤訳率、適訳率等の導入

「仮設手法」では、カウントした誤訳語数がそのまま評価結果として扱われる。だが、テストセット中の各文において選定される重要技術用語の語数は一定ではなく⁴、テストセットからどの文を評価対象文（通常 200 文）に選択するかによって、評価対象となる重要技術用語の総語数は変動する。

評価対象となる重要技術用語の総語数が異なれば、誤訳の語数が同一であっても翻訳品質は同一とは見なせないことは言うまでもない。したがって、重要技術用語の翻訳精度は、誤訳の語数ではなく、評価対象とした技術用語の総数に対する誤訳率を評価指標に用いるのが適切である。

「特許庁手法」では、重要技術用語の訳質を A（適訳語）、B（可訳語）、C（誤訳語）、D（不訳語）の 4 段階で評価し、用途に応じてそれぞれの重みづけを行うこととしている。したがって、単に誤訳率（すなわち C+D の率）のみを算出するのではなく、A～D それぞれの率を算出し、用途に応じて組み合わせて使用することとなる。

このことを踏まえ、「特許庁手法」では、A～D それぞれの比率を算出して評価結果とすることを定めた。

・ 正解率、誤訳率の算出

各重要技術用語の評価点を集計し、評価点 A～D それぞれの比率を算出する。さらに用途に応じた重みづけを行い、正解率、誤訳率として評価値に利用する。

なお、評価結果の取扱いについては、後述の「相対評価方法」（4.3.8.）にて、他の評価観点と一括して定義することとし、本項目では言及しない。

⁴ 0 語～3 語。詳しくは第 5 章 5.2.5.参照。

4.3.2.3. 訳ゆれへの対応

評価対象文において、同一の重要技術用語が文中に複数回出現する場合があるが、機械翻訳結果でこれらが常に同一の訳語で訳されるとは限らない。一つの用語が複数の異なる訳語で訳されており（「訳ゆれ」）、それぞれの訳質評価（A～D）が異なる場合、どのように扱うかをあらかじめ定義しておく必要がある。

重要技術用語数の総語数は、文中に同一の用語が複数回出現した場合も 1 語としてカウントしている。つまり、ある文において 1 種類の重要技術用語が文中に 2 回出現したとしても、1 語として扱われる。このため、この語に訳ゆれが生じ、その評点（A～D）が同一でない場合には、これを 1 つに決めなければならない。

その一案として、各評点を出現回数で除算してカウントするという方法が考えられる。つまり、文中に 2 回出現した用語の評点が A と B であれば、それぞれを 0.5 でカウントするというものである。だがこの方法は、集計作業を複雑化し、また評点の入力欄の設定も難しくなる（文中に最大何回出現するか想定困難）など、運用面での負担が大きい。訳ゆれへの対処は、文中に重要技術用語が複数回出現し、かつそれぞれの訳語が異なるという特殊な条件下に限られるものであり、こうした特殊なケースのために運用面の負担を増加させるのは効率的でない。

このため「特許庁手法」では、訳ゆれが生じており、それぞれの訳語の評点が異なる場合は、最良の評点のみを採用することとし、以下の内容の補足説明を追加した。

※ 同一の重要技術用語が 1 文において複数回出現し、異なる訳語に訳されている場合は、全ての訳語を上記基準で評価し、最も高い評価点（A＞D）のみを入力する。

なお、上記により、同一の重要技術用語が複数回出現する文において、それが 1 回のみ正確に訳された場合と、複数回正確に訳された場合とで、当該用語に与えられる評価値は同一となる。しかしながら、両者は「内容伝達レベルの評価」では確実に差がつくため、これら複数の評価観点を組み合わせて有用性を判断することで充分に対応できる。また、検索用途の評価に限れば、複数回出現する用語が 1 回でも適訳語（A 評価）に訳されていれば検索にはヒットするので、複数回適訳に訳されていた場合と同等の評価を与えるのが妥当とも考えられる。

4.3.2.4. その他の補記

「特許庁手法」では、各評価文における重要技術用語は、あらかじめ特定されている（つまり、評価者はこれを特定する必要がない）前提とした。

各評価文の重要技術用語をその都度評価者自身が選定する方式も検討したが、この方式は、各文における選定語や選定語数が評価の都度変化するため運用の難易度が上がり、評価者自身の作業負担も増加する。加えて、例えば同一システムの経時的評価を行う場合などは、同一条件で評価を行うために過去の評価対象語を記憶しておく必要も生じるなど、効率が悪い。

このため、各評価文における重要技術用語は、文をテストセットに含める際にテストセット作成者があらかじめ選定し、評価者は常にその用語を評価する方式とした。また、集計作業等の運用面や、作業ファイルの構成上の観点から、各文に対して選定する重要技術用語の上限を最大 3 語と定めた。

- ・ 重要技術用語の訳語の評価

対象文の技術分野を鑑み、各機械翻訳対象文に対してあらかじめ選定された重要技術用語（1 語～3 語）に対応する機械翻訳結果中の訳語を、以下の基準で評価する。

：

：

※ 機械翻訳対象文が重要技術用語を含まない場合は、評価対象外とする。

なお、重要技術用語の選定に関する詳細は、第 5 章 5.2.5.にて説明する。

4.3.2.5.「特許庁手法」の確定内容

上記の再検討結果を踏まえ、「特許庁手法」の定義内容を以下のとおり確定した。なお、「仮設手法」における本観点の名称は「技術用語に関する誤訳の個数」であったが、「特許庁手法」では、重要技術用語の翻訳精度を4段階（A～D）に区分し、それぞれの率を評価結果として用いる形に改めたため、観点名も「重要技術用語の翻訳精度の評価」と改めた。

表 4-6.「特許庁手法」の「(2)重要技術用語の翻訳精度の評価」

(2) 重要技術用語の翻訳精度の評価

・重要技術用語の訳語の評価

対象文の技術分野を鑑み、各機械翻訳対象文に対してあらかじめ選定された重要技術用語（1語～3語）に対応する機械翻訳結果中の訳語を、以下の基準で評価する。

A（適訳）：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）：技術用語として一般的に用いられる訳語ではないが、意味はおおむね正しい。

C（誤訳）：誤訳である。

D（不訳）：未知語、訳漏れである。

※ 同一の重要技術用語が1文において複数回出現し、異なる訳語に訳されている場合は、全ての訳語を上記基準で評価し、最も高い評価点（A＞D）のみを入力する。

※ 機械翻訳対象文が重要技術用語を含まない場合は、評価対象外とする。

・正解率、誤訳率の算出

各重要技術用語の評価点を集計し、評価点A～Dそれぞれの比率を算出する。さらに用途に応じた重みづけを行い、正解率、誤訳率として評価値に利用する。

＜技術用語のチェックは、評価対象となる各案件の属する各技術分野に精通した評価者が行う。1つの機械翻訳対象文に対し複数の機械翻訳結果を評価する場合は、その全てを同一の評価者が同時に評価する＞

4.3.3. 評価の観点及び基準：(3) 流暢さの評価

4.3.1.1.にて述べたとおり、「仮設手法」には、翻訳結果の文としての読みやすさ、理解しやすさを評価する「流暢さ」の観点が存在しない。「流暢さ」は、特許文献の機械翻訳結果を評価するうえで主要な観点ではないものの、例えば「内容の伝達レベル」や「重要技術用語の翻訳精度」等が同程度の翻訳結果の優劣を定める際の補助的な指標のひとつとして必要なものである。このため「特許庁手法」では、新たに「流暢さ」を評価する観点を追加した。

4.3.3.1. Fluency と Acceptability

「流暢さ (Fluency)」の評価は、人手評価の一指標として、主に「内容の正確さ (Adequacy)」の評価と併用される形で用いられてきた。しかしその一方で、「原文の内容を反映していなくても日本語としてスムーズでさえあれば高評価が与えられる」という性質が問題視されることもあり、事実、NTCIR-9 の特許翻訳タスクの人手評価では、この問題点に対応した新指標である Acceptability を Fluency に代えて採用している (2.1.2.3.)。

Acceptability は、「原文の意味が正しく理解できたうえで、かつ流暢であるか」を 5 段階評価するものであり、原文の意味が正確に理解できない場合は無条件で最低評価となる。したがって、流暢だが原文の意味を反映していない文が高い評価を得るという問題点は解消される。しかしその一方で、原文全体の意味を正確に反映しない限り、誤訳の程度によらず常に最低評価となってしまうため、中／低品質な翻訳結果に対する評価の解像度が低いという新たな課題が生じている。

「特許庁手法」においては、「流暢さ」はあくまで補助的な観点という位置づけであり、その用途は、例えば「内容の伝達レベルの評価」等の主要な評価観点で差がつかない場合に参照する程度のものである。このような利用方法であれば、流暢だが原文の内容を十分に反映していない機械翻訳結果は、「流暢さ」を考慮する以前の段階ですでに淘汰されるはずであり、したがって Acceptability にて問題視された点への配慮は不要である。むしろ、「内容の伝達レベルの評価」と観点が一部オーバーラップする Acceptability よりも、文としてのスムーズさのみを測定する Fluency のほうが、複数の評価観点を用途に応じて重み付けする形態の運用を想定する「特許庁手法」においては、より扱い易いと考えられる。

このため、「特許庁手法」では、「流暢さの評価」は、原文の内容を問わず機械翻訳結果の文としての理解し易さのみを評価する Fluency 方式を採用することとした。

4.3.3.2.「流暢さの評価」の評価基準

Fluency の評価基準は、第 2 章 2.1.1.3.の表 2-8 に示されたとおりの 5 段階評価である。具体的には、評点 5 が Flawless（完璧）、4 が Good（良好）、3 が Non-native（非ネイティブ）、2 が disfluent（流暢でない）、1 が Incomprehensible（理解不可能）と定義されている。

流暢さの評価に該当する評価基準としては、他にも Mu プロジェクトの用いた理解容易性（2.1.1.2.の表 2-3）や平成 20 年度特許庁調査における基準（2.1.3.1.の表 2-17）等が挙げられるが、どれも類似の基準が設定されている。

「特許庁手法」における「流暢さの評価」の基準も、これらに準拠して定めることとした。ただし、「特許庁手法」が評価対象とする機械翻訳結果は、一般文よりも翻訳の難易度の高い特許文に関するものであるため、評価基準を若干低めに設定し、以下の基準とした。

- 5：文意が明解で、人間が書いた文に近い。
 - 4：文として不自然な箇所を若干含むが、文意は明解である。
 - 3：文として不自然な箇所があり、文意がわかりにくい。
 - 2：文法規則に反する表現をかなり含む。文意がわからない。
 - 1：文として成立していない。

具体的には、評点 5 の要件を「完璧な文」とはせず、文意が明解で、ほぼ人間が書いた文に近ければ良しとした。また、「文意がわからない」ものについても、文として成立していない場合のみを評点 1 とし、文意は不明でも文として成立していれば評点 2 とし、両者を区別可能とした。

各評点に該当する典型的な機械翻訳結果のサンプルを次表 4-7 に示す。評価作業実施時は、こうしたサンプルを（本項目に限らず各項目について）作成して評価者に提示し、評価水準のずれの解消に努める必要がある。

表 4-7. 「流暢さの評価」各評点の典型例サンプル

5	文意が明解で、人間が書いた文に近い。 好ましくは、油圧シリンダーが上部のアーム組立品を操作するために使用されている一方で、電動アクチュエータは脚拡張および側面のエクステンション・アームを操作するために使用される。
4	日本語として不自然な箇所を若干含むが、文意は明解である。 装置は、3 点のカップリング構成で異なる絶頂および幅を持っている様々な道具の用途に適していて、カテゴリー1 およびカテゴリー2 道具の用途に特に適している。
3	日本語として不自然な箇所があり、文意がわかりにくい。 カップリング組立品 40 は、上部のアーム油圧シリンダー13 の操作により道具に比べて適切に角度がある。
2	日本語文法規則に反する表現をかなり含む。文意がわからない。 垂直のものがスリーブをつけることを特徴とする実施の形態において、51 および 61、ならびに拡張脚 52 と 53 角度、内部へ、この最終操作は、使用の間に、よりよく道具 91 を保持する安全な方法で、道具 91 のピン 93 および 94 を固定する役目をする。
1	日本語文として成立していない。 図. 3、クロススリーブ部材と横延長アーム、及び垂直スリーブ部材、垂直延長脚と下側フック部材を含むほぼ垂直アセンブリを含むほぼ水平立体を示す結合アセンブリの背面図であり、トラクタ下部ヒッチに接続されて示さアーム。

4.3.3.3. 「流暢さ」の定義

「流暢さ」の評価では、機械翻訳結果を独立した文として扱い、その読みやすさ、理解しやすさのみを評価する。したがって、原文の内容が反映されているか否かは考慮してはならないが、このような特殊な評価観点は、研究者を除けばあまり一般的なものとはいえない。このため「特許庁手法」では、評価者に注意を促すための注記を以下のとおり加えた。

※ 本評価は、翻訳結果のみを独立した日本語文として扱う。「文意」とは、あくまで翻訳結果において読み取れる文意を指し、原文や人手翻訳の内容と一致しているかどうかは考慮しない。

4.3.3.4. 評価の順序

「流暢さ」の評価は、前述のとおり、原文の内容が反映されているかを考慮せず、機械翻訳結果単独での文としての読みやすさ、理解し易さのみを評価する必要がある。

しかしながら、実際の評価作業においては、同時に行われる別観点の評価、具体的には「内容の伝達レベルの評価」や「重要技術用語の翻訳精度の評価」等、原文や人手翻訳の内容に照らして行う評価の過程で評価者が得た“原文の記載内容についての知識”が、流暢さを評価する際のバイアス要因となる、という問題点が、第1回委員会にて指摘された。

これを回避するためには、評価作業において、原文やその人手翻訳を参照する他の評価を行う前に、機械翻訳結果のみを参照して「流暢さ」の観点からの評価を行うことが有効である。「特許庁手法」では、このことを要件として以下のとおり補記した。

< 1つの機械翻訳対象文に対し同一の評価者が「流暢さ」の評価と他の観点からの評価の双方を担当する場合、最初に「流暢さの評価」を、原文や人手翻訳を参照せずに行う >

4.3.3.5. サンプル評価

「流暢さ」の評価は、補助的な観点として実施するものであり、評価作業の効率化の観点から、サンプル評価とするのが妥当と判断し、これを補記した。

サンプル評価の対象文数は50文と仮置きした。その妥当性については第8章 8.4.4.にて検証する。また、「流暢さ」の度合いは文長により変化すると推測されるため、サンプル中に一定数の長文を含めることも要件化した。

< 機械翻訳対象文のうちランダムに選択した50文のみを対象とするサンプリング評価を行う。一定数の長文を含める > ※

※第8章 8.4.4.に示した検証の結果、50文のサンプル評価では十分に信頼性のある相対順位は得られないことが判明した。このため、「特許庁手法」では、最終的には上記要件を見直し、「流暢さの評価」についても200文全文を対象とすることとした。詳しくは第9章 9.2.3.を参照のこと。

4.3.3.6. 「特許庁手法」の確定

上記検討結果により、「特許庁手法」における「流暢さの評価」の定義内容は、以下のとおりとなった。

表 4-8. 「特許庁手法」の「(3) 流暢さの評価」

(3) 流暢さの評価

機械翻訳結果を独立した文として見た際の、文としての自然さ、読みやすさのみを以下の 5 段階の評価基準で主観的に評価する。

- 5 : 文意が明解で、人間が書いた文に近い。
- 4 : 文として不自然な箇所を若干含むが、文意は明解である。
- 3 : 文として不自然な箇所があり、文意がわかりにくい。
- 2 : 文法規則に反する表現をかなり含む。文意がわからない。
- 1 : 文として成立していない。

※ 本評価は、翻訳結果のみを独立した日本語文として扱う。「文意」とは、あくまで翻訳結果において読み取れる文意を指し、原文や人手翻訳の内容と一致しているかどうかは考慮しない。

<機械翻訳対象文のうちランダムに選択した 50 文のみを対象とするサンプリング評価を行う。一定数の長文を含める>

<1つの機械翻訳対象文に対し同一の評価者が「流暢さ」の評価と他の観点からの評価の双方を担当する場合、最初に「流暢さの評価」を、原文や人手翻訳を参照せずに行う>

<1つの機械翻訳対象文に対し複数の機械翻訳結果を評価する場合は、その全てを同一の評価者が同時に評価する>

4.3.4. 評価の観点及び基準：(4) チェックリストに基づく評価①「基本方針」

「仮設手法」の「(3) チェックリストに基づく評価」は、その冒頭で以下のとおり評価の手法を定義している。

表 4-9. 「仮設手法」の「(3) チェックリストに基づく評価」定義部分

機械翻訳特有の課題を対象言語に応じて洗い出したチェックリストを作成し、チェックリストに該当する翻訳ミスの個数をカウントする。
--

「特許庁手法」への「チェックリストに基づく評価」の採否を検討するにあたり、まずはこの評価観点の特徴及び用途について再確認した。

4.3.4.1. 「チェックリストに基づく評価」の特徴

「チェックリストに基づく評価」とは、上表 4-9 の定義のとおり、機械翻訳結果に含まれる翻訳ミスを 1 件ずつ抽出し、該当するチェック項目にカウントしていく形式の評価を指す。評価結果からは、特定の種類の翻訳ミスが翻訳結果にどれだけ含まれていたかが分かる。翻訳ミスの数を数え上げる方式であるため、客観的な評価観点といえる。

評価作業においては、評価者は、機械翻訳結果から全ての翻訳ミスを検出し、場合によってはその原因を特定する必要がある。このため、他の評価観点における評価者の要件である「当該技術分野に関する専門知識」に加えて、原語読解力も必要となる。また、機械翻訳エンジンに関するある程度の知識も有することが望ましい。

機械翻訳結果から全ての翻訳ミスを検出する作業は、他の評価観点の作業に比べ、評価者の作業負担が極めて大きい。特に長文においては、多数の翻訳ミスが複合的に生じるため、作業に要する時間や作業の難易度は著しく増大する。このため、評価対象とする文の選定や文数の設定には十分な配慮が必要である。

4.3.4.2. 「チェックリストに基づく評価」の用途

「チェックリストに基づく評価」の主な用途として、評価における客観性の担保のために他の評価観点と併用するという用途と、翻訳ミスのカウント結果からシステムの誤訳の傾向や弱点を把握し、システムの性能改善のためのフィードバックに利用するという用途が挙げられる。

前者は、「基本的な考え方」（添付資料 1-1）の「5. 客観性の高い評価」でも言及されているものである。しかしながら、「チェックリストに基づく評価」は、機械翻訳結果中の翻訳ミスを文意への影響の軽重を問わずにカウントする方式であるため、評価結果が必ずしも機械翻訳結果の有用性と比例するとは限らない。このため、これを客観性の担保目的で

相対評価や絶対評価に含めると、機械翻訳結果の実務への有用性を測るという「特許庁手法」本来の目的をむしろ損ねる危険性がある。

この問題への対応策として、翻訳ミスを重大なものと軽微なものに分けてカウントする方式も検討した。だがこの方式は、重大・軽微の判断を評価者の主観で行うこととなるため、客観的評価という本評価観点の特質が失われ、客観性の担保に適さなくなる。

これらを踏まえ、「特許庁手法」では、客観性の担保のために「チェックリストに基づく評価」を相対評価や絶対評価の一観点として使用することは適切ではないと判断した。「特許庁手法」は、「内容の伝達レベル」における複数の評価者による重複評価や事前の意識合わせ、評価サンプルの充実、そして、比較的客観性のある「重要技術用語の翻訳精度の評価」を評価観点に含めること等で、相対評価や絶対評価の客観性は十分に担保されており、前記リスクを冒してまで「チェックリストに基づく評価」をこれらに含める必要はないと考える。

ただし、「チェックリストに基づく評価」のもう一つの用途である、システムの性能改善のためのフィードバック利用に関しては、「特許庁手法」に含まれる他の評価観点の評価結果では目的を十分に達成できない。例えばあるシステムで「内容の伝達レベル」の評価結果が悪かったとしても、主にどのような翻訳ミスが原因でそのような評価となったのかは評価結果から特定できないため、システムの具体的な弱点や傾向は把握できない。

このため「特許庁手法」では、フィードバック用途専用の評価観点として「チェックリストに基づく評価」を採用する方針とした。

4.3.4.3. 「特許庁手法」の「チェックリストに基づく評価」の全体構成

「仮設手法」における「チェックリストに基づく評価」の全体構成は、冒頭に前掲（表 4-9）の定義部分があり、続いて＜課題の例＞として、あらゆる翻訳言語において共通と見なせる課題の例と、特定の翻訳言語に特有の課題の例として、中日、韓日などいくつかの翻訳言語対における特有課題とが示されている（添付資料 1-2 参照）。チェックリスト評価の実施時は、これらの例を参考に評価対象に応じた課題を改めて洗い出し、チェック項目とする想定となっている。

機械翻訳において、翻訳言語に応じた特有の課題が多数存在することは、本調査にて実施した「英日機械翻訳の現状分析（第 3 章 3.2.）」の結果からも明らかである。こうした課題は他の翻訳言語ではチェック項目として意味をなさないものも多く、「仮設手法」にて言語共通の課題と翻訳言語別の課題とを区別して扱っているのは適切である。

また、チェック項目を定めず、評価実施時に対象言語に応じて洗い出す想定としている点も、翻訳言語別の課題に関しては、あらゆる翻訳言語についてあらかじめ全ての課題を洗い出すことは実質的に不可能であり、妥当な対処といえる。ただし、言語共通のチェック項目については、あらかじめこれを確立しておくことも可能であり、またそのほうが運用上効率的である。

このため「特許庁手法」では、言語共通の課題、すなわち言語共通のチェック項目をあらかじめ定義しておくこととした。また、翻訳言語別のチェック項目についても、英日機械翻訳に関しては、「英日機械翻訳の現状分析」にて特有の課題や留意点を十分に把握しているため、その内容に基づいてチェック項目を定義した。

なお、これらチェック項目は、必ずしも常時全ての項目をチェック対象とする必要はないと思われるため、言語共通のチェック項目を【必須項目】と【任意項目】とに区分し、後者については評価対象や評価の目的等に応じて評価実施時に任意に選択可能とした。また、翻訳言語別の項目に関しては、原則として全て任意項目とし、【翻訳言語別任意項目】と名づけた。

評価対象文数については、「仮設手法」では他の評価観点と同時に評価する前提であったため特に定義されていなかったが、「特許庁手法」では本観点はフィードバック利用のために他の観点と切り離して単独で行われる前提としており、改めてこれを定義する必要がある。ただし、フィードバックのための評価は、必ずしも一定以上の文数の評価が必要という性質のものではなく、例えば「数式、化学式部分の認識誤り」という課題に対し、数式や化学式を含む文を数件評価する、というような運用もあり得る。このため、「特許庁手法」では評価対象文数は特定の文数に定めず、任意に設定可能とした。

下表 4-10 に、上記検討内容を踏まえて作成した「特許庁手法」の「チェックリストに基づく評価」の全体構成を示す。なお、全体構成中の【必須項目】、【任意項目】、【翻訳言語別任意項目 [英日]】それぞれの詳細については、次節以降で詳述する。

表 4-10. 「特許庁手法」の「(4) チェックリストに基づく評価」全体構成

(4) チェックリストに基づく評価

選定した機械評価対象文の機械翻訳結果に含まれる翻訳ミスの個数を、以下に示すチェックリストに基づきカウントする。

チェックリストは、以下に示す【必須項目】をベースとし、必要に応じて【任意項目】及び【翻訳言語別任意項目】から任意に選択した項目をこれに追加する。

評価対象文数は任意に設定してよい。

【必須項目】

：

【任意項目】

※ 任意項目は、評価の目的や作業負担等を鑑み、必要な項目のみを使用する。

：

【翻訳言語別任意項目】

※ 翻訳言語別任意項目は、評価の目的や作業負担等を鑑み、必要な項目のみを使用する。

[英日]

：

＜評価は、機械翻訳対象文に用いられる言語の読解力を備え、かつ評価対象となる各案件の属する各技術分野に精通した評価者が行う。1 つの機械翻訳対象文に対し複数の機械翻訳結果を評価する場合は、その全てを同一の評価者が同時に評価する＞

4.3.5. 評価の観点及び基準：(4) チェックリストに基づく評価②【必須項目】

「特許庁手法」では、「チェックリストに基づく評価」にてチェックリストに含めるべき機械翻訳の課題、すなわちチェック項目を、あらゆる翻訳言語において共通の課題と、翻訳言語に応じた課題とに区分して定義し、前者はさらに、常にチェックリストに含める【必須項目】と、任意選択が可能な【任意項目】とに区分することとした。本節では、まずは言語共通課題の【必須項目】についての検討結果を記す。

4.3.5.1. 「仮設手法」のチェックリスト項目の試用

「仮設手法」では、チェックリスト項目とすべき機械翻訳の課題が＜課題の例＞として示されており、前半部が言語共通課題、後半部が翻訳言語に応じた課題となっている。

下表 4-11 に、「仮設手法」の＜課題の例＞の前半部、すなわち「言語共通の課題」部分を示す。

表 4-11. 「仮設手法」の「(3) チェックリストに基づく評価」言語共通の課題部分

＜課題の例＞

構文解析に関する誤り

- ・主語、動詞、目的語（他動詞の場合）の組み合わせが誤っている。
- ・主語、動詞、目的語（他動詞の場合）の対応が把握できない。
- ・動詞が複数ある場合に、副詞（句）の係り受け先が誤っている。
- ・名詞が複数ある場合に、形容詞（句）の係り受け先が誤っている。
- ・重複した係り受け（複数の形容詞（句）が1つの名詞に係り受けする場合、及び、複数の副詞（句）が1つの動詞に係り受けする場合）の訳出を誤っている。
- ・並列構造の取り違い

文法事項に関する誤り

- ・品詞を取り違えている。
- ・時制が誤っている。
- ・修飾語と被修飾語の位置関係が不適切。

語句、フレーズに関する誤り

- ・専門用語に対して一般的な訳語となっている。
- ・異なる技術分野の用語に訳されている。
- ・多義語が正しく訳されていない。
- ・未知語が含まれている。
- ・訳抜けがある。
- ・特許文献に定型的な表現（～を特徴とする〇〇、請求項〇記載の～、等の表現）に対応できていない。

- ・発明の名称について体言止めができていない。
- ・出願人名が正しく訳されていない。

記号等の処理に関する誤り

- ・数式、化学式の部分の認識誤り（途中で分断して訳出してしまっている。）
- ・箇条書き表現（";", ":", "(a)", "a.", "(1)", "1.", 等）が適切に処理されていない。
- ・本体部分とカッコ書き部分の位置関係が不適切。
- ・数字部分と単位部分の位置関係が不適切。

「特許庁手法」の検討にあたり、まずは上表に示した「仮設手法」の言語共通課題部分をそのままチェックリスト項目として試用し、これらが実用可能であるかを検証した。その結果、以下の3つの問題が判明した。

- ① どの項目にも該当しない翻訳ミスが多発し、これをカウントできない。
- ② 翻訳ミスが複数の項目に該当する場合があります、その際の扱いが決められていない。
- ③ 係り受け誤りの条件が具体的すぎ、どれにも該当しないケースが多発する。

例えば、下表 4-12 に示す機械翻訳結果サンプルに対し、「仮設手法」のチェックリストを試用してみる。

表 4-12. 機械翻訳結果サンプル

原 文	To drive two toggle levers (29, 29') of a press, in particular a high frequency cutting press with at least 400 strokes per minute, two separate but mutually synchronized drives (101, 101') are provided to ensure a uniform, non-wobbling motion of the tool carrier (5).
	プレス機、特に1分あたり400ストローク以上の高周波切断プレス機、の2つのトグルレバー（29、29'）を駆動するため、分離しているが相互に同期する2つの駆動部（101、101'）を提供し、これによって工具キャリア（5）の均一で安定した動作を保証する。
機械翻訳	特に、プレスの2つのトグルレバー（29' および29）を駆動するために、少なくとも400ストローク／分を備えた高周波切断プレス、2、分かれている、しかし相互に駆動（101および101'）を同期させた、工具キャリア（5）の一定の動揺しない動作を保証するために供給される。

この機械翻訳結果における翻訳ミスとして、以下の10件が挙げられる。

- #1：29 と 29' とが逆の順序で出力されている。
- #2：「in particular」は「high frequency cutting press」に係るべきだが「To drive...」に係っている。
- #3：「a high frequency...per minute」は「a press」に係るべきだが誤っている。

- #4: 「with」を「を備えた」と訳しているが、「の」等と訳すほうが適切。
- #5: 「two separate...」の「two」は「drives」に係るべきだが、誤っている。
- #6: 「separate」は「drives」に係るべきだが、誤っている。
- #7: 「mutually synchronized」は「drives」を修飾する形容詞だが、動詞として扱われている。
- #8: 「drives」が「駆動」と訳されているが、動作ではなく物体を指しており「駆動部」等と訳すほうが適切。
- #9: 「101'」のアポストロフィ (') が閉じカッコ (」) になっている。
- #10: 「are provided」は「drives」に係るべきだが、誤っている。

これら 10 件の翻訳ミスを「仮設手法」のチェックリストと照合すると、下表 4-13 のようになる。

表 4-13 翻訳ミスとチェックリスト項目の照合結果

No.	翻訳ミス内容	チェックリスト該当項目
#1	29 と 29' の順序誤り	該当なし
#2	「in particular」係り受け誤り	該当なし
#3	「a high frequency...」係り受け誤り	該当なし
#4	「with」訳語不適	「多義語が正しく訳されていない」
#5	「two」係り受け誤り	「重複した係り受けの訳出誤り」および「修飾語と被修飾語の位置関係が不適切」？
#6	「separate」係り受け誤り	「重複した係り受けの訳出誤り」または「並列構造の取り違え」？
#7	「mutually synchronized」品詞誤り	「品詞を取り違えている」及び「並列構造の取り違え」
#8	「drives」訳語不適	「多義語が正しく訳されていない」
#9	「101'」のアポストロフィ扱い不適	該当なし
#10	「are provided」係り受け誤り	「主語、動詞、目的語の組み合わせ誤り」または「主語、動詞、目的語の対応が把握できない」

翻訳ミス#1 は、原文の「(29, 29')」が訳文では「(29' および 29)」と逆の順序になってしまったものであるが、「仮設手法」のチェックリスト項目にはこの翻訳ミスに該当するものが存在せず、カウントすることができない。チェックリスト項目には「本体部分とカッコ書き部分の位置関係が不適切」というものがあるが、本件はカッコ書き内部における問題であるため該当しない。なお翻訳ミス#9 にもほぼ同様のことが言える。

翻訳ミス#2 と#3 は、どちらも係り受け誤りに属するが、チェックリストに列挙された条件のいずれにも合致せず、こちらもカウントできない。#5、#6 も係り受け誤りであり、これらは「重複した係り受けの訳出を誤っている」という項目に該当すると見なせるが、前者はさらに「修飾語 (two) と被修飾語 (drives) の位置関係が不適切」、後者は「並列構

造 (separate but mutually synchronized) の取り違え」にも該当するとも考えられ、この場合、双方にカウントすべきか否かが判断できない。

なお、上記機械翻訳結果はルールベース方式の機械翻訳 (RBMT) によるものであったが、統計翻訳方式 (SMT) の機械翻訳結果においても、例えば「provided : 用意されている」と訳されるべき箇所が「用意されてい」で切れていたり、原文の「101」のアポストロフィが訳文では別の語に付されていたりと、やはり「仮設手法」がカバーしていない翻訳ミスが数多く検出された。

こうした試行の結果、「仮設手法」のチェックリスト項目は、そのままの形では実用不可能と判明した。したがって「特許庁手法」にて、これを実用可能な形に修正する必要がある。

4.3.5.2. 【必須項目】の概念の導入

「特許庁手法」のチェックリスト項目では、「仮設手法」における下記 3 つの問題点を解決する必要がある。

- ① どの項目にも該当しない翻訳ミスが多発し、これをカウントできない。
- ② 翻訳ミスが複数の項目に該当する場合があります、その際の扱いが決められていない。
- ③ 係り受け誤りの条件が具体的すぎ、どれにも該当しないケースが多発する。

このうち①については、チェックリストに「その他の翻訳ミス」という項目を新設することでとりあえずは解決する。だが、「仮設手法」にこの項目を加えるだけでは、性質の異なる多数の翻訳ミスがここに分類されてしまい、「機械翻訳システムの弱点や誤訳の傾向を把握する」というチェックリスト評価の目的を十分に果たすことができない。このため、各種機械翻訳において生じる可能性のある主要な翻訳ミスは、全て項目化してチェックリストに網羅しておく必要がある。

機械翻訳における主要な翻訳ミスには、原文に存在する語に対応する訳語が翻訳結果に存在しない「語の欠落 (A)」、逆に原文に存在しない語が翻訳結果に混入する「不要語の混入 (B)」、原文の用語が翻訳されずそのまま出力される「未知語 (C)」、多品詞語が誤った品詞に解釈される「品詞の誤り (D)」、語間や文節間における「係り受け誤り (E)」、用語の訳語が不適である「語の誤訳 (F)」の 6 種類が挙げられる。「特許庁手法」では、これら 6 つの主要な翻訳ミスに「その他の翻訳ミス (G)」を加えた 7 項目 (A~G) をチェックリストの【必須項目】とし、全ての翻訳ミスが確実にカウントされるようにした。

さらに、【必須項目】のカウントは 1 つの翻訳ミスに対して 1 回のみと定め、複数の項目に該当する場合は最上位の 1 項目のみにカウントすることとした。例えば「light : 光」という語が「軽い」と誤訳されていた場合、これは品詞誤り (D) でもあり語の誤訳 (F) で

もあるが、カウントは上位の D のみに行うこととなる。こうすることで、「仮設手法」の問題点の②「翻訳ミスが複数の項目に該当する場合があります、その際の扱いが決められていない」が解決し、さらに、【必須項目】のカウント合計値にて当該機械翻訳結果における翻訳ミスの総数が把握可能となる。

問題点③「係り受け誤りの条件が具体的すぎ、どれにも該当しないケースが多発する」ことへの対応策としては、主要な係り受け誤りの条件を「仮設手法」の粒度で網羅する方法と、係り受け誤りの条件設定を緩めて個々のカバー範囲を広げる方法とが考えられる。

しかしながら、前者に関しては、「仮設手法」の粒度で係り受けの種類や因果関係を網羅していくとなると、チェック項目数が膨大となることは避けられない。例えば「仮設手法」の「動詞が複数ある場合に、副詞（句）の係り受け先が誤っている」という条件については、「動詞」「複数」「副詞（句）」の各部分を「名詞」「単数」「形容詞（句）」等に展開して掛け合わせた全ての条件をチェック項目化することとなる。評価者の作業負担や評価の難易度を考えると、このようなチェックリストは実用性に欠ける。

したがって、選択肢は実質的には後者、すなわち「係り受け誤りの条件設定を緩めて個々のカバー範囲を広げる」方法のみといえる。「特許庁手法」は、こちらの方法を前提に検討を進めた。

その具体的な内容として、検討当初は、係り受け誤りの条件別分類という方針自体は維持し、条件設定をより大きな括り、具体的には「文骨格レベルの誤り」「節レベルの誤り」「句レベルの誤り」「語レベルの誤り」に区分する案を試してみた。しかし、実際の機械翻訳結果においては、これらの誤りが複合的に発生していて解析が困難を極めるものや、特に SMT 方式など文法規則に依存しない方式の機械翻訳結果においては、文法的な分類が困難なケースも多く、この案も実用困難と判明した。

このため「特許庁手法」では、【必須項目】において係り受け誤りを条件別に分類すること自体を断念し、全ての係り受け誤りをまとめてカウントする方式とした。そして、これを補うため、係り受け誤りのなかでも特に意味把握において重大なもの（主語・述語・目的語の関係性の誤り等）や機械翻訳において特に課題とされるもの（並列表現に関する誤り等）は、別途【任意項目】として設定することとした。

4.3.5.3. 翻訳ミスのカウント単位

「チェックリストに基づく評価」を行うにあたり、あらかじめ翻訳ミスのカウント単位を明確に定義しておく必要がある。「特許庁手法」では、翻訳ミスのカウント単位について、「語単位、係り受け関係単位」を原則とした。

〔語単位のカウント〕

「語単位のカウント」における「語」は、複合語や慣用表現を含むものとした。つまり複合語や慣用表現は1語とみなされる。具体的には、例えば「liquid crystal display : 液晶ディスプレイ」という複合語が、例えば「液体水晶展示」と誤訳されていた場合、単語単位では liquid、crystal、display の3語とも不適訳であるが、翻訳ミスは複合語単位でカウントするため、「語の誤訳」へのカウントは3ではなく1となる。慣用表現についても同様に、「in addition to : ～に加えて」という慣用表現が「～のために加算の中で」等と誤訳されている場合、in、addition、to の3語ではなく、in addition to が翻訳ミスのカウント単位となる。

複合語については、多数の語から構成されるような場合、どこまでを1語と見なすかは定義が難しいが、判断の参考として、以下に一例を示す。

「脂肪アルコールを含む組成物」

⇒ fatty alcohol containing composition は1語とみなす。

⇒ composition containing fatty alcohol は composition、containing、fatty alcohol に分ける。

⇒ fatty alcohol contained composition は1語とみなす。

⇒ fatty alcohol contained in composition は fatty alcohol、containing、in、composition に分ける。

〔係り受け関係単位のカウント〕

続いて「係り受け関係単位のカウント」という定義の解釈について、下記具体例にて説明する。

原 文	Preferably, a contact member or members is mounted to the container spaced from and facing the contact member or members carried by the upright restraint.
翻訳結果	好ましくは、間隔を置いて配置されて、接点部材に向かっているコンテナまたは直立した制限により担持される部材に、接点部材または部材は、載置される。

上の例における係り受け誤りを「係り受け関係単位」でカウントすると、以下の4箇所となる。

- #1 : 「mounted to : 載置される」の目的語は「container : コンテナ」のみだが、翻訳結果においては container と (2 回目の) 「members : 部材」の双方が目的語となっている。
- #2 : 「spaced from : 間隔を置いて」の目的語は (2 回目の) contact member or members だが、翻訳結果においては目的語が存在しない。
- #3 : 「facing : 向かっている」も (2 回目の) contact member or members が目的語だが、翻訳結果では「contact member : 接点部材」のみに係っている。
- #4 : 「carried : 担持される」も (2 回目の) contact member or members を修飾しているが、翻訳結果では「members : 部材」のみに係っている。

これら 4 つの係り受け誤りは、全て「2 回目の contact member or members が並列と解釈されていない」ことに起因しており、したがって 1 つの翻訳ミスであるという考え方も可能である。しかしながら、こうして個々の係り受け誤りの根本原因を辿っていく作業は非常に難易度が高く、また SMT 方式等では不可能な場合も多い。このため、「特許庁手法」のチェックリストにおける翻訳ミスのカウント単位は、あくまで語単位・係り受け単位とし、上例のように 1 つの語 (contact member or members) が複数の係り受け関係を持つ場合は、それぞれの係り受け関係をカウント対象とすることとした。

なお、カウントは係り受け関係にある語のどちら側からカウントしても問題ない。上に示した例は係る側からのカウントだが、係られる側からカウントしたとしても、以下のとおり、カウント数は同一となる。

- #1 : (2 回目の) 「members : 部材」が誤って「mounted to : 載置される」の目的語となっている。
- #2 : (2 回目の) contact member or members が「spaced from : 間隔を置いて」の目的語となっていない。
- #3 : (2 回目の) 「members : 部材」が「facing : 向かっている」の目的語となっていない。
- #4 : (2 回目の) 「contact member : 接点部材」が「carried : 担持される」を受けていない。

「特許庁手法」には、このカウント単位に関する定義を以下のとおり補記した。なお、詳しい解説は、評価作業時に評価者に配布する評価手順書 (7.1.3.1.参照) にて行う。

翻訳ミスのカウントは、語単位 (複合語や慣用表現は一語と見なす)、係り受け関係単位でのカウントを原則とする。

〔重要技術用語の取り扱い〕

「チェックリストに基づく評価」は、その評価結果を相対評価や絶対評価のために他の観点の評価結果と組み合わせて使用することは想定していないが、評価作業自体は他の観点の評価と同時に実施することもある。その際は、「重要技術用語の翻訳精度の評価」にて評価済みの語も改めてチェックリスト評価の対象とするが、この点に誤解のないよう、これを明示する以下の文言を補記に加えておく。

重要技術用語として評価した語も再度評価対象とする。

4.3.5.4. 重複カウントの禁止

「特許庁手法」では、【必須項目】のカウントは1つの翻訳ミスに対して1回のみと定め、1つの翻訳ミスが複数のチェック項目に該当する場合は最も上位の項目のみにカウントすることとした。

例えば「B.品詞の誤り」は、常に「D.語の誤訳」を併発し、多くの場合「C.係り受け誤り」も併発する（例えば名詞「well：井戸」が副詞「良く」と品詞誤りされた場合、語の誤訳にも該当し、係り受けも誤ったものとなる可能性が高い）が、このような場合、カウントは最上位の「B.品詞の誤り」のみにカウントすればよい。

ただし、1語において発生する翻訳ミスは、常に1つだけとは限らない。したがって、1語において複数の翻訳ミスが発生している場合は、それぞれを個別にカウントする必要がある。それは例えば以下のような場合を指す。

原 文	This permits the lower edges of the windscreens to slide over the upper surface of the slide bars 9,10 without hindrance from the brake bars 25,26.
翻訳結果	これは、フロントガラスの下方端が制動棒 25、26 からスライド・バー9、10 の上面の無事上を滑るのを許容する。

この例における **without hindrance** は「妨げられることなく」と訳すのが適当だが、翻訳結果では「無事に」となっており、まずは「語の誤訳」としてカウントする必要がある。さらに、原文からはこの語は「slide：滑る」に係るべきものであり、語の誤訳があるにせよ少なくとも「無事に滑る」と訳されなければならないところ、翻訳結果ではそうっておらず、（おそらく）**upper surface** に係っている。したがって「係り受けの誤り」としても、再度カウントする必要がある。

先に示した「well：井戸」の事例との違いは、「well：井戸」においては、上位の誤り（品詞誤り）が常に下位の誤り（語の誤訳、係り受け誤り）の直接の原因となっているのに対し、本例は、一方の事例（語の誤訳）が他方の事例（係り受け誤り）の原因となっておら

ず、それぞれが独立して発生している点にある。通常、1 語に対し複数の翻訳ミスが生じる場合、本例と同様「係り受け誤り」と「誤の誤訳」とが併発するケースが多い。

重複カウントの禁止に関して、「特許庁手法」には以下のとおり補記を追加した。

1 個の翻訳ミスは【必須項目】中の 1 項目のみにカウントする。翻訳ミスが複数の項目に該当する場合は、順序が上位の項目のみにカウントする。
1 語に対して複数の異なる翻訳ミスが生じている場合は、個別にカウントする。

4.3.5.5. 翻訳ミスの範囲

意味的には誤りとはいえないが、訳文がたどたどしく流暢さを欠く表現が機械翻訳結果にはしばしば見られる。「特許庁手法」においては、こうした表現は翻訳ミスとはみなさないこととした。その旨、以下のとおり明記する。

日本語として読みにくいが意味的・文法的に誤解を与える懸念の少ない語や語順、文構造は、「流暢さ」の範疇であり、ここでは翻訳ミスと見なさない。

4.3.5.6. 「特許庁手法」の確定内容

「特許庁手法」の「チェックリストに基づく評価」の【必須項目】部分の定義内容は、上記検討を経て以下のとおり確定した。

表 4-14. 「特許庁手法」の「(4) チェックリストに基づく評価」【必須項目】部分

【必須項目】

機械翻訳結果に含まれる全ての翻訳ミスを、下記 A～G の該当する項目 1 箇所にカウントする。1 つの翻訳ミスが下記項目の複数に該当する場合は、上位の項目のみにカウントする。

- A. 語の欠落
- B. 不要語の混入
- C. 未知語
- D. 品詞の誤り
- E. 係り受け誤り
- F. 語の誤訳
- G. その他の翻訳ミス

[チェック手順]

- ・ 機械翻訳結果に含まれる全ての翻訳ミスを【必須項目】に分類し、個数をカウントする。
- ・ 翻訳ミスのカウントは、語単位（複合語や慣用表現は一語と見なす）、係り受け関係単位でのカウントを原則とする。重要技術用語として評価した語も再度評価対象とする。
- ・ 1 個の翻訳ミスは【必須項目】中の 1 項目のみにカウントする。翻訳ミスが複数の項目に該当する場合は、順序が上位の項目のみにカウントする。
- ・ 1 語に対して複数の異なる翻訳ミスが生じている場合は、個別にカウントする。
- ・ 日本語として読みにくいが意味的・文法的に誤解を与える懸念の少ない語や語順、文構造は、「流暢さ」の範疇であり、ここでは翻訳ミスと見なさない。

<評価は、機械翻訳対象文に用いられる言語の読解力を備え、かつ評価対象となる各案件の属する各技術分野に精通した評価者が行う。1 つの機械翻訳対象文に対し複数の機械翻訳結果を評価する場合は、その全てを同一の評価者が同時に評価する>

4.3.5.7. 【必須項目】各項目の定義

【必須項目】の各チェック項目（A～G）について、カウント対象となる翻訳ミスの種類を、具体例⁵を添えて説明する。

[A. 語の欠落]

原文中の語に対応する訳語が、機械翻訳結果中に存在しないケースを指す。なお、人手翻訳は原文中にない語が補足されている可能性がある⁶ため、欠落か否かの判断は原文と機械翻訳結果とを対比して行う必要がある。

原 文	Physiologically, the voltage- difference in the myocardial cells of the heart should be close to 0, because the ventricles were just activated and had contracted (in the QRS).		
人手翻訳	生理学的に、(QRSにおいて)心室がちょうど活性化され、収縮したので、心筋細胞の 電圧差 は0近くにあるべきである。		
機械翻訳	生理的に、 電圧 —心臓の心筋細胞において、室がちょうど起動して、収縮した(QRS)ので、約0でなければならない。		
翻訳ミス	「Voltage-difference：電圧差」の「difference：差」の部分の訳語が欠落している。	該当項目	A

[B. 不要語の混入]

原文中に対応する語が存在しない不要な訳語が機械翻訳結果に混入しているケースを指す。

原 文	Behind the front panel a predetermined distance can be a reflective panel or other surface.		
人手翻訳	前面パネルの後ろの予め定められた 距離 には反射パネルまたはその他の表面があってもよい。		
機械翻訳	フロントパネルの後ろでは、予定された 車間の距離 は、反射的なパネルまたは他の表面であるかもしれない。		
翻訳ミス	原文・人手翻訳に存在しない「車間の」という語が混入している。	該当項目	B

⁵ 各具体例における原文や翻訳結果は、説明上の必要に応じて加工しており、実際の機械翻訳結果そのものではない場合がある。

⁶ 原文と人手翻訳が意味的に対応していない対訳文はテストセットには採用されないが、文意に影響しない範囲での適切な言い換えは許容されており（5.2.3.2.）、その結果、原文にない語が補足される等の可能性がある。

<不要語の混入と見なさないケース>

ただし、機械翻訳システムが必要に応じ補足／省略したとみなせる語は、文意に悪影響がない限り不問とする。

原 文	This invention relates generally to the field of mechanisms used to temporarily connect an implement to a tractor, and more particularly to mechanisms known as three point hitch systems or assemblies.		
人手翻訳	本発明は、器具をトラクタに一時的に接続するために使用される機構の分野全般に関し、より詳細には、3点ヒッチシステム又は3点ヒッチ組立体として知られている機構に関する。		
機械翻訳	本発明は、通常、一時的に道具をトラクタに接続するために用いる機構の分野に関する。より詳しくは、 本願発明は 三点ヒッチ・システムまたはアセンブリとして公知の機構に関する。		
翻訳ミス	機械翻訳結果には、原文中に対応する語が存在しない「本願発明」という語が存在するが、これはエンジンが訳文を2文に分割した際の補足であり、不要語の混入とはみなさない。	該当項目	—

[C. 未知語]

原文の一部が機械翻訳されず原語のまま出力されているケースを指す。ただし、意図的に原語のまま出力していると考えられる語（例：ping packet⇒ping パケット）は対象外である。

原 文	The side of the container opposite the column 11 is provided with complementary pairs of paddles 21, 22 tiltably mounted to a post 23 at the side of the container.		
人手翻訳	柱11の反対側のコンテナの側面には、角度を変えることができるようにコンテナの側面のポスト23に取り付けられた補完的なパドルの対21、22が設けられている。		
機械翻訳	容器の側面に t i l t a b l y にポスト23に載置されるパドル21、22の相補的な対を、柱11の反対側の容器の側面は、備えている。		
翻訳ミス	Tiltably が未知語となっている。	該当項目	C

[D. 品詞の誤り]

原文中の語が、誤った品詞として翻訳されているケースが対象となる。

原 文	Fig. 4 presents detail of an embodiment of one jaw of the transferring element.		
人手翻訳	図4は、移送構成要素の1つのジョーの実施形態の詳細を 示す 図である。		
機械翻訳	図伝達要素の1顎の形態の4 プレゼント の詳細。		
翻訳ミス	動詞 present (示す) が名詞 (プレゼント) と誤って解釈されている。	該当項目	D

原 文	an upper hook and two lower hooks, said lower hooks disposed on two extension legs telescopically retained by two vertical sleeve members;		
人手翻訳	上部フック及び２つの下部フックであって、 前記 下部フックが、２つの垂直スリーブ部材により入れ子状に保持された２つの拡張脚部に配置されており、		
機械翻訳	上部のフックと２下フックは、下のフックは伸縮縦の二重スリーブ部材が保持する２つの拡張脚上に配置されたと 述べた ；		
翻訳ミス	形容詞 said （前記）が動詞（述べた）と誤って解釈されている。	該当項目	D

<複合語、イディオム>

複数の語から成る複合語（の一部）やイディオムがそのように解釈されていない場合も品詞誤りと見なし、複合語・イディオム単位で１つの語と扱ったうえで本項目にカウントする。

原 文	[0024] A heat shield 7 is interposed between the upper face of heating body 3 and housing 1, the heat shield being maintained on the heating body 3 by fastening means.		
人手翻訳	【００２４】熱シールド７は 加熱ボディー ３の上面とハウジング１との間に挿入され、該熱シールドは固定手段によって加熱ボディー３上に維持される。		
機械翻訳	【0024】熱シールド７は 本体 ３と、ハウジング１を 加熱する の上面との間に介在され、熱シールドは締結手段によって発熱体３上に維持されている。		
翻訳ミス	Heating body（加熱ボディー）が複合語として扱われず、heating が誤って動詞（加熱する）に訳されている。こうした場合、heating body の単位で品詞誤りとみなす。	該当項目	D

<自動詞・他動詞の誤りは「語の誤訳（F）」とする>

自動詞と他動詞の誤りは、本項目ではなく「F. 語の誤訳」に該当するものとする。

原 文	None of the outer and inner flanges 723a, 723b can overlap one another when the inner layer 719b is disposed in the outer layer 719a.		
人手翻訳	内側層 719b が外側層 719a 内に配置される際に、どの外側及び内側フランジ 723a、723b も互いに 重なり合 わないようにすることができる。		
機械翻訳	内部層 719b が外層 719a に配置されるとき、外側そして内側のフランジ 723a、723b のいずれも、互いに 重ね合わせる ことができない。		
翻訳ミス	原文の overlap は自動詞（重なり合う）で訳すべきだが、機械翻訳結果は他動詞（重ね合わせる）で訳されている。このような場合は、品詞の誤り（D）ではなく語の誤訳（F）にカウントする。	該当項目	F

〔E. 係り受け誤り〕

機械翻訳結果において各語間の係り受けが正しくない場合、及び、文法規則に反するような語の配置については、この項目にカウントする。カウントは、係り受け関係単位で行う（4.3.5.3.参照）。

原 文	Not infrequently the pulley must be removed from the pin by means of a very quick and simple operation, for example so as to replace it with a different pulley.		
人手翻訳	例えばそれを異なる滑車に置き換えるために、非常に急速で簡単な操作によって該ピンから滑車を取り外さなくてはならないことは 稀ではない 。		
機械翻訳	まれに 、例えばそれを異なるプーリに置き換えるように、プーリは、あまり速いおよび簡単な運営によって、ピンから取り除かれては ならない 。		
翻訳ミス	「Not infrequently」は「時々」の意味だが、Not が infrequently ではなく must be removed に係り、誤った意味になっている。	該当項目	E

原 文	advantageously, the core 5 is made of nylon, and the layer 6 of conductive metal is for example made of silver.		
人手翻訳	有利には、コア 5 はナイロンでできており、導電性金属の層 6 は、 例えば 銀でできている。		
機械翻訳	有利に、コア 5 はナイロンからなる。また、導電性金属の層 6 は銀からなって、 例えば ある。		
翻訳ミス	「for example：例えば」は「made：なる」に係るべきものだが、そうっていない(「is：ある」に係っている)。	該当項目	E

＜「係り受け誤り」と「語の誤訳（F）」の区別＞

要素の組合せは正しいが、その関係性が誤っている、もしくは助詞の欠落等により関係性が不明瞭な場合も「係り受け誤り」とみなす。

原 文	The first and second ear straps are bonded to the mask body by a first bonding part and a second bonding part , respectively, which are located away from each other in a first direction respectively.		
人手翻訳	第 1 及び第 2 の耳掛け部は、それぞれ第 1 の方向に沿って離れている 第 1 の接合部および第 2 の接合部によって マスク本体部に接合されている。		
機械翻訳	第 1 及び第 2 の耳のストラップは、 第 1 の接合部と第 2 の接合部は 、それぞれ、第 1 方向に互いに離隔されるマスク本体に接合されている。		
翻訳ミス	「by a first bonding part～：第 1 の接合部～」は「bonded to：接合されている」に係るものであり、機械翻訳結果でもその点は正しいが、by を「によって」ではなく「は」と訳しているため、関係性が正しくない。この翻訳ミスは「by の誤訳（F）」と見なすことも可能だが、より上位の本項目にカウントする。	該当項目	E

<係り受け誤りとみなさないケース>

例えば能動文から受動文への言い換えなどによって主語・動詞・目的語等の要素が入れ替わり、結果として原文と異なる係り受けとなっている場合でも、原文における各要素の相互関係が正しく把握できる適切な係り受けであれば、誤りとは見なさない。

原 文	[0021] The strength of the flexible hose can be increased if the pressure support elements comprise a metallic wire.		
人手翻訳	〔 〇 〇 2 1 〕 可撓性ホースの 強度は 、圧力支持部品が金属ワイヤーを含む場合に 増加される 。		
機械翻訳	[0021] 圧力支持体素子をメタリックワイヤから構成する場合、可撓性ホースの 強度を増加させることができる 。		
翻訳ミス	原文では主語である strength が機械翻訳結果では目的語となっているが、意味するところは同じであるため、係り受け誤りとはみなさない (pressure support elements についても同様)。	該当項目	—

<機械翻訳エンジンによるカッコの使用>

機械翻訳エンジンによっては、係り受け関係を本文中で表現せず、カッコ書きを用いて表現する場合がある。このような場合でも、係り受け関係が正しく把握される限り、エンジンがカッコ書きを用いたことを理由に翻訳ミスとみなす必要はない。

原 文	It is preferred that the container is of frame construction, preferably metallic frame construction.		
人手翻訳	コンテナはフレーム構造、好ましくは金属フレーム構造であると好ましい。		
機械翻訳	コンテナがフレーム構造（好ましくは金属的なフレーム構造）であることが好ましい。		
翻訳ミス	Preferably 以下の frame construction への係り受けが、本文ではなくカッコ書きで表現されているが、正しい係り受け関係が維持されているため翻訳ミスとはみなさない。	該当項目	—

[F. 語の誤訳]

原文中の語が適訳語（同義語として一般的に用いられる用語の範疇の語）に訳されていない場合、この項目にカウントする。カウントは語単位を原則とし、複合名詞（例：liquid crystal display）やイディオム（例：among others）等は、全体を1語として扱う。

原 文	The arm usually forms an angle of 90 with the base .		
人手翻訳	アームは通常、 基部 と90度の角度を形成する。		
機械翻訳	アームは、通常90の角度を 塩基 により形成する。		
翻訳ミス	「Base：基部」が「塩基」と、異なる技術分野の訳語に誤訳されている。	該当項目	F

原 文	The contact member advantageously comprises a flexible or resilient contact layer for contacting the adjacent item in the container.		
人手翻訳	接触部材は、コンテナの中で隣り合う 品物 と接触するための可撓性接触層または弾力的な接触層を含むと有利である。		
機械翻訳	接触部材は、有利には、容器内の隣接する 項目 と接触するための可撓性又は弾性接触層を含む。		
翻訳ミス	Item は「項目」ではなく「品物」等の意味。	該当項目	F

一文で同じ語が複数回誤訳されている場合は、1回のみカウントすることとする。

なお、「重要技術用語の翻訳精度の評価」を同時に実施した場合も、全ての「重要技術用語」が再度評価対象となる。原則、「重要技術用語の翻訳精度の評価」にてB（可訳語）～C（誤訳）と評価された語はこの項目に該当する（※「品詞の誤り」の場合を除く）。

原 文	A further slide bar 62 is fixed at the bottom of the container adjacent openable door panel 5.		
人手翻訳	コンテナの基部において、開くことができる扉の板5と隣り合って別の スライドバー 62が固定されている。		
機械翻訳	更なる すべり棒 62は、開けられるドアーパネル5に隣接してコンテナの底で修正される。		
翻訳ミス	「Slide bar：すべり棒」は技術的に同義かどうか微妙だが、意味は分かるため「重要技術用語の翻訳精度の評価」では「B. 可訳語」に該当するが、「チェックリストに基づく評価」ではこうした訳語も「語の誤訳」と見なす。	該当項目	F

〔G. その他の翻訳ミス〕

上記 A～F のいずれの項目にも該当しない翻訳ミスをここにカウントする。例えば以下のようなケースが考えられる。

原 文	For example in the embodiment shown in the drawings, doors 5 and 7 pivot on respective support posts to effectively open respective corners of the container, thereby to enhance access to the ends of the container for loading or unloading of the container.		
人手翻訳	たとえば、図面に示された実施態様において、扉 5 および 7 はそれぞれの支柱を軸として回転してコンテナのそれぞれの角を効果的に開き、それによって、コンテナの荷積みまたは荷降ろしのためのコンテナの端への出し入れを容易にする。		
機械翻訳	ドローイングの実施例でネットワーク・イメージ用のネットワーク共有を作成する場合は、ルートのインストール先の UNC パスを半角 39 文字未満にしてください、コンテナの事実上開いているそれぞれのコーナへのそれぞれの支柱のドア 5 と 7 旋回軸; その結果、コンテナの装填または取り降ろしのためのコンテナの端部への機能アップアクセスに。		
翻訳ミス	機械翻訳結果中に原文と無関係な文が混入しているが、内容的に「B. 不要語の混入」の範疇を超えており、システムエラーの可能性もあるため、「その他の翻訳ミス」に分類した。	該当項目	G

原 文	To drive two toggle levers (29, 29') of a press,...		
人手翻訳	プレス機の 2 つのトグルレバー (29、29') を駆動するために、...		
機械翻訳	プレスの 2 つのトグルレバー (2 9' および 2 9) を駆動するために、		
翻訳ミス	カッコ中の付番 29 と 29' の順序が逆になっている。	該当項目	G

〔備考〕

翻訳ミスが A～F のどの項目に該当するか特定困難な場合は、最も可能性の高いと思われる項目にカウントすることとする。判断がつかない場合は「G. その他の翻訳ミス」にカウントする。

4.3.6. 評価の観点及び基準：(4) チェックリストに基づく評価③【任意項目】

「特許庁手法」では、「チェックリストに基づく評価」における言語共通課題を【必須項目】と【任意項目】とに区分した。本節では【任意項目】について説明する。

4.3.6.1. 【任意項目】設定理由

「仮設手法」の＜課題の例＞には、必須項目・任意項目の区別はないが、内容的には、例えば「品詞を取り違えている」「未知語が含まれている」「異なる技術分野の用語に訳されている」等、「特許庁手法」の【必須項目】とほぼ同一の切り口のものと、これとは全く異なる切り口のものとが混在している。

後者は、例えば「特許文献に定型的な表現に対応できていない」といった、翻訳ミスの結果ではなく原因による分類であったり、「並列構造の取り違い」や「数字部分と単位部分の位置関係が不適切」など、翻訳ミスの中でも機械翻訳が特に苦手とする条件に限定したものであるため、重複カウントを認めない【必須項目】に組み込むことが難しい。しかしながら、これらの観点は、機械翻訳システムの弱点や翻訳ミスの傾向を把握するためには非常に重要である。

【任意項目】は、こうした“【必須項目】に組み込めないが評価観点として重要”な項目をカバーするために設けたものである。「特許庁手法」では、下記 12 項目を【任意項目】に設定した。

- Z-00・訳語のゆれ
- Z-01・不適切な文の分断
- Z-02・主語・述語・目的語（他動詞の場合）の相互関係に関する翻訳ミス
- Z-03・並列表現に関する翻訳ミス
- Z-04・発明の名称など、名詞句を文として扱った翻訳ミス
- Z-05・特許特有の用語や表現に関する翻訳ミス
- Z-06・特許特有の記号や文献番号表記、付番に関する翻訳ミス
- Z-07・箇条書き表現に関する翻訳ミス
- Z-08・原文のカッコ書きに起因する翻訳ミス
- Z-09・数式、化学式に関する翻訳ミス
- Z-10・数字と単位に関する翻訳ミス
- Z-11・数値範囲に関する翻訳ミス

これら 12 項目には、「仮設手法」に定義されていた項目に加え、「英日機械翻訳の現状分析」の過程で検出された課題や第 1 回委員会での指摘内容等を踏まえて「特許庁手法」にて新規に設定したものも含まれている。

4.3.6.2.【任意項目】のカウント基準

【必須項目】では、各項目の合計をもって機械翻訳結果に含まれる翻訳ミスの総数とするため、個々の翻訳ミスを必ずカウントし、かつ重複カウントを禁止している。これに対し【任意項目】は、【必須項目】にカウントされた翻訳ミスを、異なる切り口や特定のケースに限定して再カウントするものであるため、必ず【必須項目】との重複カウントとなる。また、各【任意項目】の切り口は多様であり、1つの翻訳ミスが複数の任意項目に該当することも考えられる。このため、【任意項目】の各項目間の重複カウントも許容される。

なお、【任意項目】は、【必須項目】と異なり、全ての翻訳ミスが必ずカウントされる性質のものではなく、したがって各項目のカウント値を合計することに大きな意味はない。このため、チェックリストにどの項目を含めるかは、評価の目的等に応じて任意に決められるものとした。

4.3.6.3. 「特許庁手法」の確定

上記を踏まえ、「特許庁手法」の「チェックリストに基づく評価」における【任意項目】の定義は、以下のとおりとした。

表 4-15. 「特許庁手法」の「(4) チェックリストに基づく評価」【任意項目】

(4) チェックリストに基づく評価

：

【任意項目】

【必須項目】にカウントした翻訳ミスについて、特に下記項目にあてはまる場合はここにもカウントする。1 つの翻訳ミスが下記項目の複数に該当する場合は、該当する全ての項目にカウントする。

Z-00・訳語のゆれ

Z-01・不適切な文の分断

Z-02・主語・述語・目的語（他動詞の場合）の相互関係に関する翻訳ミス

Z-03・並列表現に関する翻訳ミス

Z-04・発明の名称など、名詞句を文として扱った翻訳ミス

Z-05・特許特有の用語や表現に関する翻訳ミス

Z-06・特許特有の記号や文献番号表記、付番に関する翻訳ミス

Z-07・箇条書き表現に関する翻訳ミス

Z-08・原文のカッコ書きに起因する翻訳ミス

Z-09・数式、化学式に関する翻訳ミス

Z-10・数字と単位に関する翻訳ミス

Z-11・数値範囲に関する翻訳ミス

※【任意項目】は、評価の目的や作業負担等を鑑み、必要な項目のみを使用する。

[チェック手順]

：

- ・ 各翻訳ミスは、【必須項目】にカウントした後、【任意項目】の各項目と照合し、該当する全ての項目にもカウントする。

4.3.6.7.【任意項目】各項目の定義

【任意項目】の各項目（Z-00～Z-11）について、それぞれの項目がカウント対象とする翻訳ミスの種類等の詳細を、具体例とともに解説する。

[Z-00. 訳語のゆれ]

原文中に同じ用語が複数回出現し、機械翻訳結果においてそれらが異なる訳語で訳されていると、読解に悪影響を及ぼす可能性がある。項目 Z-00 では、原文中に同一の用語が複数回出現し、それらの訳語が統一されていないことで読解に悪影響を及ぼす場合をカウント対象とする。したがって、前置詞や冠詞をはじめ、読解にあたって訳語が統一されている必要のない語は対象外となる。

評価の際、各訳語の適・不適を考慮する必要はない。つまり、訳ゆれが生じていれば、たとえ双方の訳語が適訳であっても（例：silicon が「ケイ素」「シリコン」と訳ゆれ）、本項目のカウント対象となる。カウントは（出現回数単位でなく）原語単位で行うものとする。上例の場合、silicon が対象となり、カウントは 1 となる。

なお、本項目に該当する翻訳ミスは、上例のように訳ゆれの双方が適訳である場合、【必須項目】A～F のいずれにも該当しない。このような場合は【必須項目】は「G. その他の翻訳ミス」にカウントする。なお、訳ゆれ的一方または双方が A～F のいずれかに該当する場合は、上位項目を優先し、G へはカウントしない。

以下に、本 Z-00 項目に該当する訳ゆれの具体例を示す。

原 文	[0025] Housing 1 contains a water reservoir 8 being composed of an upper part 9 and a lower part which are assembled on one another, heat shield 7 defining the lower part of water reservoir 8.		
人手翻訳	【 0 0 2 5 】ハウジング 1 は、互いに組み立てられる上側部分 9 および下側部分から構成される 水リザーバ 8 を包含し、熱シールド 7 が 水リザーバ 8 の下側部分を画定する。		
機械翻訳	[0025]はハウジング 1 上部 9 と互いに組み立てられ下部で構成されている 貯水槽 8 が含まれている、熱シールド 7 は 水 の下部 貯水池 8 を定義する。		
翻訳ミス	原文中 2 度出現する water reservoir について、機械翻訳結果では 1 回目が「貯水槽」、2 回目が「水の／貯水池」と異なる訳語に訳されており、訳語のゆれに相当。 ※2 回目の water reservoir は複合語として扱われていないため必須チェック項目の D にも該当する。	該当項目	Z-00

<対象語が語単位で一致しない場合>

同一の語であっても、一方が単語、他方が複合語の一部であったり、それぞれが異なる複合語に含まれていたりする場合がある。このような場合は、同一の訳語に訳されないことと読解に支障をきたすかどうかで判断する。

原 文	It will be appreciated that a given application service is dedicated to a predetermined application .		
人手翻訳	特定の アプリケーション ・サービスは、所定の アプリケーション に専用されることが理解されよう。		
機械翻訳	特定の アプリケーション ・サービスが予定された 塗布 に捧げられるのが理解される。		
翻訳ミス	原文に 2 度出現する application は、1 回目は application service という複合語、2 回目は application という単語であるが、文意から同一の事項を指しており、同一の訳語に訳すべきものである。 ※「application：塗布」は「F. 語の誤訳」にも該当する。	該当項目	Z-00

[Z-01. 不適切な文の分断]

例えば入力文が長すぎて処理できない場合や、記号等に用いられたピリオドを誤って文末と判断してしまった場合など、不適切な文の分断が検出された場合、ここにカウントする。

ただし、例えば要素列挙時や、適切な箇所での意図的な分割など、正当な理由が認められる場合は対象外とする。

原 文	[0014] FIG. 3 shows a cross-sectional view of the fixing device according to FIG. 1 along the cross-sectional plane I-I of FIG. 1 .		
人手翻訳	【0014】図3は、 図1 の横断平面 I-I に沿う 図1 に従う固定装置の横断面図を示す。		
機械翻訳	【0014】 図 。図3は、図1に係る定着装置の断面図を示す。 断面平面図の I-I に沿って 1 。 1 。		
翻訳ミス	FIG. 1 のピリオドが文末と解釈され、文が分断されている。 2 箇所において発生しているため、Z-01 は 2 回カウントする。 ※FIG.1 はカウント単位では 1 語とみなされるが、機械翻訳結果においてはそう扱われていないため項目 D にも該当する。また、後述の Z-06（特許特有の記号に関する翻訳ミス）にも該当する。	該当項目	Z-01×2

[Z-02. 主語・述語・目的語（他動詞の場合）の相互関係に関する翻訳ミス]

原文における骨格構造、すなわち主語・述語・目的語（他動詞の場合）の相互関係が正しく保たれていない場合、ここにカウントする。

原 文	The ping packets is expected to return a response within the TTL = 1.		
人手翻訳	ピング・パケットは 、T T L = 1 内に応答を返すことが 期待される 。		
機械翻訳	ピング・パケット T T L = 1 の中で応答を返す 予想する 。		
翻訳ミス	原文の骨格構造は「主語＝ピングパケット、動詞＝期待される」だが、機械翻訳結果は、ピングパケットが主語かどうか不明瞭であり、かつ動詞「予想する」が（意味が誤っていることを不問としても）能動態となっているため、骨格構造を維持できていない。 ※【必須項目】においては「ping packet と is expected」の係り受け誤り（E）と、expect の誤訳（F）の 2 つの翻訳ミスとしてカウントが必要だが、Z-02 へのカウントは 1 回でよい。	該当項目	Z-02

<便宜的な主語等の扱い>

It is preferred that XXX など、it や there 等の代名詞が便宜的な主語として用いられている場合は、実質的な主語（XXX）を評価対象とする。

原 文	It is preferred that one or more contact members are height adjustable with respect to the container.		
人手翻訳	1 つ以上の接触部材がコンテナに対して高さ調節可能であると好ましい。		
機械翻訳	それは 、一つ以上の接触部材は、容器に対して高さ調節可能であることが好ましい。		
翻訳ミス	It は便宜的な主語であって翻訳文は that 以下を主語として扱うべきであるが、機械翻訳結果では it をそのまま主語として扱い「それは」と訳しているため、本項目に該当するとみなした。 ※翻訳すべきでない It を訳しているという翻訳ミスは、「G. その他の翻訳ミス」とした。	該当項目	Z-02

<能動・受動の言い換え等の扱い>

なお、例えば能動文から受動文への言い換えなどにより主語・動詞・目的語が入れ替わるような場合も、これらの「相互関係」が正しく把握される限り、本項目の対象とはならない。

原 文	Energy absorption means (21) is provided at each end for absorbing energy transmitted by the ropes from the impacting vehicle.		
人手翻訳	衝突車両からロープによって伝達されるエネルギーを吸収するために、各端部には エネルギー吸収手段（２１）が設けられる。		
機械翻訳	各端のときにロープで衝撃付与乗り物からエネルギー透過を吸収するのに エネルギー吸収手段(21)を設ける。		
翻訳ミス	原文は「エネルギー吸収手段が設けられる」という受動文。機械翻訳結果ではこれを「エネルギー吸収手段を設ける」と能動態に言い換えているが、主語と動詞の関係性は維持されているので翻訳ミスとはみなさない。	該当項目	—

<主語、述語、目的語の誤訳>

主語や述語が誤訳されて正しい意味が把握できなくても、文の骨格構造が維持されていれば本項目には該当しない。

原 文	The tongue engages inside a corresponding seat of the pulley once the latter has been inserted on the pinion.		
人手翻訳	該 トング は、滑車がピニオンに挿入されると滑車の対応する シート の内側に 係合する。		
機械翻訳	後者はピニオンに挿入された後に 舌が 滑車の対応 座席 の内側 を行っています。		
翻訳ミス	tongue が「舌」、engage が「行っている」と誤訳されているため文意は把握困難だが、文の骨格構造は正しいため本項目には該当しない。	該当項目	—

＜文全体が名詞句である場合＞

本項目は原文が文であるもののみを対象とする。したがって、「発明の名称」や「特許請求の範囲」など、対象文全体が名詞句である場合は対象外となる。「特許請求の範囲」などにおいて、こうした名詞句を機械翻訳エンジンが文形式で翻訳することがあるが、この場合も対象外とする。

原 文	4. Device (10) according to claim 3, wherein said element (60) which can be operated by the user is slidable along said cavity (14).		
人手翻訳	4. 請求項3に記載の装置（10）において、ユーザにより操作され得る前記エレメント（60）は前記空洞（14）に沿ってスライド可能である、装置。		
機械翻訳	4.請求項3に記載のデバイス（10）において、ユーザにより作動されることができる前記素子（60）は、前記空腔（14）に沿って摺動可能である。		
翻訳ミス	機械翻訳結果は意図的に文形式で翻訳されており主語、述語が存在するが、原文が名詞句であるため、本項目の対象としない。	該当項目	—

[Z-03. 並列表現に関する翻訳ミス]

原文に含まれる並列関係が機械翻訳結果において正しく認識されず、その構造が崩れていたり、双方に係るべき修飾語が一方のみに係っていたりする場合にカウントする。並列表現は、付番の並列や語の並列から句の並列、節の並列まで、種類を問わず対象とする。

原 文	A second duct includes a damper or valve which may be detachable.		
人手翻訳	第2のダクトは、取り外し可能な ダンパー または バルブ を備える。		
機械翻訳	別のダクトは、 ダンパー または取り外せる バルブ を有する。		
翻訳ミス	damper と valve が並列関係にあり、双方を which 以下が修飾しているが、機械翻訳ではそう見なされず、which が valve のみを修飾している。 ※which 節の係り受け誤り（E）にも該当する。	該当項目	Z-03

なお、本項目の評価観点は「並列関係が保たれているか」という一点のみであり、それさえ保たれていれば、他種の翻訳ミス（例えば語の誤訳）が発生していても本項目のカウント対象とはしない。

原 文	A second duct includes a damper or valve which may be detachable.		
人手翻訳	第2のダクトは、取り外し可能な ダンパー または バルブ を備える。		
機械翻訳	第2ダクトは、取り外せる ピアノダンパー または バルブ を備える。		
翻訳ミス	Damper がピアノダンパーと誤訳されているが、valve との並列関係は which 節の係り受けを含め保たれているので、本項目には該当しない。	該当項目	—

<並列表現における繰り返し語の省略>

並列表現において繰り返し部分が省略されるケースがあり、これに適切に対応できているかどうか本項目の重要な評価観点となる。

原 文	As an additive and modificative process , after a trench portion is formed in the inorganic insulating film, …		
人手翻訳	追加的、修正的プロセス として、無機絶縁膜にトレンチ部を形成した後、トレンチ部の内壁面に、付着膜として、…		
機械翻訳	添加剤および変更するプロセス として、溝部分が例えば無機絶縁膜(CVD 薄膜)の中で形成された後、…		
翻訳ミス	additive と modificative はともに process を修飾する形容詞であり、additive の後の process が省略されている。機械翻訳結果はこの省略に対応できずに additive を「添加剤」と名詞解釈しているため、並列表現に関する翻訳ミスとみなす。 ※additive は品詞誤り (D) にも該当。	該当項目	Z-03

繰り返し語の省略は、動詞や助動詞等においても発生する。

原 文	At the unloading location, the brake system can be disengaged and the paddles released.		
人手翻訳	荷降ろしの場所において、ブレーキシステムの係合を解き、パドルを 離すことができる 。		
機械翻訳	取り降ろし所在地では、ブレーキシステムは、解放されていて リリースされた パドルであるかもしれない。		
翻訳ミス	末尾の the paddles released は正しくは the paddles can be released であり can be が省略されているが、そう解釈されず、released が後置修飾の形容詞と判断されている。 ※released の品詞誤り (D) にもカウント。なお上例では disengaged も形容詞に誤解釈されている。	該当項目	Z-03

<複数の並列関係の複合>

複数の並列関係が複合するケースにおいては、個々の並列関係ごとに評価対象とする。

原 文	In addition, solar and otherwise powered, lighted signs are common.		
人手翻訳	加えて、太陽電池または別様で電力供給される点灯式標識も一般的である。		
機械翻訳	太陽で、動力がそうでなければ供給されて、さらに、明かりのついたサインは一般的である。		
翻訳ミス	並列表現は①solar と otherwise powered、そして②solar and otherwise powered と lighted の 2 箇所。①②とも sign を修飾しているが、機械翻訳結果では、どちらも正しく把握できないため①、②とも本項目に該当。したがって Z-03 は 2 カウントとなる。 ※他に、otherwise が品詞誤り (D)、In addition と solar がそれぞれ係り受け誤り (E)、powered が語の誤訳 (F) にそれぞれ該当する。	該当項目	Z-03×2

[Z-04. 発明の名称など、名詞句を文として扱った翻訳ミス]

原文全体が名詞句である場合において、機械翻訳結果にてこれを誤って文と解釈している場合、ここにカウントする。また、通常の文においても、文の一部を構成する名詞句が文として誤解釈されていればカウント対象とする。

原 文	TRANSFERRING ELEMENT OF A PUSHER MECHANISM OF A GLASS FORMING MACHINE		
人手翻訳	ガラス成形機のプッシャ機構の 移送構成要素		
機械翻訳	ガラスフォーミングマシンのプッシャー機構の 原理を移す 。		
翻訳ミス	名詞句を文として訳している。 ※品詞誤り（D）にも必ず該当する。	該当項目	Z-03

<意図的な文形式での翻訳>

機械翻訳結果において名詞句が意図的に文として訳されている場合は対象外となる。

原 文	2. The integrated circuit chip of claim 1, wherein the at least one functional domain includes an RF macro.		
人手翻訳	【請求項 2】前記少なくとも 1 つの機能領域が R F マクロを含む、請求項 1 に記載の集積回路チップ。		
機械翻訳	【請求項 2】請求項 1 に記載の集積回路チップは、少なくとも 1 つの機能性領域は、R F マクロを含むものである。		
翻訳ミス	機械翻訳結果は請求項の典型的な名詞句パターンを意図的に文形式で翻訳していると判断されるため、そのこと自体は翻訳ミスとみなさない。	該当項目	—

[Z-05. 特許特有の用語や表現に関する翻訳ミス]

英語特許文献であれば「detailed description：詳細な説明」や「preferred embodiment：好適な実施形態」、「incorporated by reference：参考として援用された」など、通常の文では滅多に用いられないが、特許文献においては頻繁に用いられる用語や表現に関する翻訳ミスをカウントする。技術用語全般は対象としない。

原 文	One skilled in the art will understand that numerous changes and variations are possible without departing from the spirit of the invention.		
人手翻訳	本発明の趣旨から逸脱することなく多数の変更および変形が可能であることを当業者は理解するだろう。		
機械翻訳	技術に熟練している 1、多数の変更および変化が発明のスピリットから外れずに可能であると理解するだろう。		
翻訳ミス	「one skilled in the art：当業者」「spirit of the invention：発明の趣旨」は特許特有の表現であり、機械翻訳結果はこれらを正しく訳せていないため本項目にカウントが必要。 ※どちらも 語の誤訳 (F) にも該当する。	該当項目	Z-05×2

なお、「発明の名称」における代名詞 same や「特許請求の範囲」等における wherein 等、語としては一般語であっても特許文献にて特有、もしくは頻出かつ典型的な用いられ方をする語については、本項目の対象とする。

[Z-06. 特許特有の記号や番号表記、付番に関する翻訳ミス]

具体的には、図表番号 (Fig.1 や FIG1)、文献番号 (US.Pat.No.) 等の記号や番号表記に関する翻訳ミスをカウントする。下例のような要素への付番も対象となる。

原 文	The locking teeth 52a,b are connected to the arms 54a,b by means of two inclined segments 62a,b.		
人手翻訳	ロッキング歯 52 a、b は 2 つの傾斜セグメント 6 2 a、b によってアーム 5 4 a、b に結合されている。		
機械翻訳	係止歯 52a は、b は 2 つの傾斜部セグメント 62a、b によって腕部 54a、b に接続されている。		
翻訳ミス	Locking teeth の付番 52a,b が正しく解釈されず、b が単独の要素として扱われている。 ※係り受け誤り (E) のほか、a と b は並列表現でもあるため Z-03 (並列表現に関する翻訳ミス) へのカウントも必要。	該当項目	Z-06

[Z-07. 箇条書き表現に関する翻訳ミス]

評価対象文が箇条書き文に該当し、その箇条書き表現部分において翻訳ミスが発生している場合、この項目にカウントする。特許文献で多用される段落番号や請求項番号も、箇条書きの一種と見なし、本項目の対象とする。

原 文	[0034] The crystal consequently obtained was cut perpendicularly to the direction of growth.		
人手翻訳	[0 0 3 4] その結果得られた結晶を成長方向に対して垂直に切断した。		
機械翻訳	[0034] は、その結果得られた結晶を成長方向に垂直に切断した。		
翻訳ミス	段落番号[0034]が文の一部として扱われてしまっている。 ※段落番号を名詞と誤解釈しているため、品詞の誤り (D) にも該当。	該当項目	Z-07

<請求項番号に関する注意>

請求項番号に関しては、原文の箇条書き番号が「1.」「2.」等の一般的な項番であるのに対して、人手翻訳ではこれが【請求項1】【請求項2】等と訳される場合がある。だが、機械翻訳エンジンは、冒頭の箇条書き番号が請求項番号であるか一般的な項番であるかは、文献上の記載位置により判断することが多いため、単文単位で機械翻訳を行う「特許庁手法」では、これらを判別することは難しい。したがって、請求項番号に関しては、原文が一般的な項番である場合、機械翻訳結果がこれを一般的な項番として翻訳しても、翻訳ミスとは見なさないこととする。

原 文	5. Testing method according to claim 1 wherein the sample of ore to be tested is between 2 and 10 kg in weight.		
人手翻訳	【請求項5】 試験される前記鉱石試料は、重量が2 k g から1 0 k g の間である、請求項1に記載の試験方法。		
機械翻訳	5. テストされる鉱石のサンプルが重さ2～10kgであることを特徴とする請求項1に記載の試験方法。		
翻訳ミス	箇条書き番号が人手翻訳では【請求項5】となっているのに対し機械翻訳結果は通常の項番のままだが、原文は一般的な項番であり、これに忠実に訳されているため、翻訳ミスとは見なさない。	該当項目	—

[Z-08. 原文のカッコ書きに起因する翻訳ミス]

原文がカッコ書きを含み、そのカッコ書き表現に起因した翻訳ミスが発生している場合にカウントする。具体的には、カッコで括られた範囲が原文と異なる場合や、カッコ書き部分と他の部分との係り受けが不正な場合等が該当し、カッコ内の語における係り受け誤りや誤訳など、カッコ書きとの因果関係が認められない翻訳ミスは対象としない。また、原文にないカッコ書き表現が機械翻訳結果にて用いられることがあるが、本基準の対象はあくまで原文のカッコ書きに起因する翻訳ミスであるため、これらは対象外である。

原 文	The force 40 exerted by the vertical stream 30 of air is higher than the one in the case of known structures with horizontal blowing as the air stream 30 flows vertically around the product 2 and acts on a larger area 44 of the product 2 (see fig. 7).		
人手翻訳	空気の鉛直な流れ 30 によって働かされる力 40 は、水平方向の吹き込みを伴う公知の構造の場合の力よりも高く、これは気流 30 が製品 2 の周りを鉛直に流れて製品 2 のより大きな範囲 44 に作用するためである (図 7 を参照)。		
機械翻訳	空気の垂直流 30 によって及ぼされる力 40 は、水平の (図 30 を参照して、オリジナル 2 のより大きな領域 44 で製品 2 作用周りに上下方向に流れる空気流として送風で知られている構造の場合、1 つ以上である 7)。		
翻訳ミス	カッコ書き表現 (see fig. 7) の単位が維持されず、本文の内容がカッコ中に混入してしまっている。 ※係り受け誤り (E) と特許特有記号 (Z-06) にもカウントする。	該当項目	Z-08

[Z-09. 数式、化学式に関する翻訳ミス]

評価対象文が数式や化学式を含み、その数式や化学式において翻訳ミス (式が途中で分断されている等) が発生している場合、ここにカウントする。

原 文	Given a triangle with vertices A, B, and C, the pixel intensity at point D can be obtained as: $D = A \cdot w_A / T_{sub.abc} + B \cdot w_B / T_{sub.abc} + C \cdot w_C / T_{sub.abc}$ (2)		
人手翻訳	頂点 A、B、および C を有する三角形を考慮すると、点 D での画素輝度は、以下の数式 3 で取得されることが可能である。 $D = A \times w_A / T_{abc} + B \times w_B / T_{abc} + C \times w_C / T_{abc}$		
機械翻訳	頂点 A、B、および C、点 D でのピクセル強度を得ることができるとともに、三角形所与： $D = A \cdot w_A \text{ の } / T_{sub.abc} + B \cdot w_B \text{ は } / T_{sub.abc} + C \cdot w_C / T_{sub.abc}$ (2)		
翻訳ミス	数式部分がそう認識されず、助詞が混入している。 ※助詞は不要語の混入ともみなせるが、数式は一語として扱うべきところ通常の文として扱った結果であるため、品詞の誤り (D) とするのがより適切。	該当項目	Z-09

[Z-10. 数値単位に関する翻訳ミス]

数字とそれに付随する単位（例：meter、minute、kg、% by weight）に関する翻訳ミスをカウントする。数字と単位が分断されている場合だけでなく、単位が誤訳されている場合等も対象となる。

原 文	Wafers of such seed crystal grown to the approximate level of 3 inches are now available on the market.		
人手翻訳	こうして約3インチレベルに成長させた種結晶のウェーハが現在市販されている。		
機械翻訳	3 つ の大体のレベルまで育てられたそのような種晶のウェファースは現在、 何インチ も市販で入手できる。		
翻訳ミス	原文の 3 inches が数字+単位として扱われていない。 ※3 inches で一語とみなすべきであり、これが一語として扱われていないため品詞誤り（D）に該当。	該当項目	Z-10

[Z-11. 数値範囲に関する翻訳ミス]

「from NNN to NNN」や「less than NNN」等、数値範囲を正しく処理できていない場合、ここにカウントする。

原 文	Advantageously, the fraction of weight of the electrically conductive filament 4 compared with the weight of the yarn 1 according to the present implementation method is between 5% and 50%, and preferably between 30% and 40% , for example approximately 35%.		
人手翻訳	有利には、本実施方法による糸1の重量に対する導電性フィラメント4の重量の割合は、5%から50%、好ましくは 30%から40% 、例えば約35%である。		
機械翻訳	有利には、糸1の重量と比較して電氣的に導電性フィラメント4の重量の本実装方法に係る分率は5%～50%、 30%、40% 、例えば、約35%の間であることが好ましい。		
翻訳ミス	数値範囲は「between 5% and 50%」と「between 30% and 40%」の2箇所あるが、機械翻訳では後者が数値範囲であることが表現できていない。 ※係り受け誤り（E）に該当する。	該当項目	Z-11

<数値単位に関する翻訳ミスとの区分>

なお、数値範囲は数値単位表現を伴うことが多いが、本項目で対象とするのは数値範囲の部分のみであり、数値単位に関する翻訳ミスは項目 Z-10（「数字と単位に関する翻訳ミス」）のみにカウントすればよい。

原 文	The recommended blink rate is around 55-60 times a minute.		
人手翻訳	推奨される点滅速度は1分間に約55～60回である。		
機械翻訳	推奨点滅速度は 55-60 倍分程度である。		
翻訳ミス	機械翻訳結果は数値範囲「55-60」を正しく訳しているので本項目の対象とはならないが、数値単位 times を「倍」と誤訳しているので Z-10 にはカウントが必要。 ※語の誤訳（F）にも該当する。	該当項目	Z-10

4.3.7. 評価の観点及び基準：(4) チェックリストに基づく評価④【翻訳言語別任意項目】

「仮設手法」では、「(3) チェックリストに基づく評価」の後半部分にて、中日、韓日、日中、日英それぞれの翻訳言語対による機械翻訳にて特に課題となる点や留意点を、「翻訳言語に応じた項目」として2～7つ例示している（掲載省略。添付資料 1-2 参照）。

「特許庁手法」もこれに倣い、「チェックリストに基づく評価」に【翻訳言語別任意項目】を設け、ここに翻訳言語に応じたチェック項目をまとめることとした。本調査においては、第3章 3.2.「英日機械翻訳の特有課題・留意点及び対応策」にて調査した英日機械翻訳における主要な特有課題 16 種を、[英日] 用のチェック項目としてここに反映した。

4.3.7.1. 【翻訳言語別任意項目】のカウント基準

【翻訳言語別任意項目】は、言語共通の【任意項目】と同様、【必須項目】との重複カウント項目とし、さらに【翻訳言語別任意項目】の各項目間の重複カウントや【任意項目】との重複カウントを可能とした。また、チェックリストに含める項目についても、【任意項目】と同様、評価の目的等に応じて任意に選択できるものとした。

4.3.7.2. 「特許庁手法」の確定内容

「特許庁手法」における【翻訳言語別任意項目 [英日]】の内容を下表 4-16 に示す。

表 4-16. 「特許庁手法」の「(4) チェックリストに基づく評価」【翻訳言語別任意項目 [英日]】

(4) チェックリストに基づく評価
：
【翻訳言語別任意項目】
【必須項目】及び【任意項目】にカウントした翻訳ミスについて、特に下記項目にあてはまる場合はここにもカウントする。1 つの翻訳ミスが下記項目の複数に該当する場合は、該当する全ての項目にカウントする。
※【翻訳言語別任意項目】は、評価の目的や作業負担等を鑑み、必要な項目のみを使用する。
[英日]
EJ-01・文中のピリオドを適切に処理していない
EJ-02・「:」や「;」を用いた文を適切に処理していない
EJ-03・重要な単数・複数情報が維持されていない
EJ-04・単数形・複数形併記表現を適切に処理していない
EJ-05・序数詞表現が正しく解釈されていない
EJ-06・慣用句の中間への語の挿入を適切に処理していない
EJ-07・倒置文を適切に翻訳していない
EJ-08・名詞と代名詞の関係を適切に表現できていない
EJ-09・慣用表現とみなすべき場合／そうでない場合を適切に翻訳できていない
EJ-10・前置詞に対し適切な訳語を選択できていない
EJ-11・接頭辞、接尾辞を伴う語を適切に処理していない
EJ-12・スラッシュを用いた並列表現を適切に処理していない
EJ-13・therein, therefor, wherein, 等を適切に処理していない
EJ-14・代名詞 same、the same を適切に処理していない（【発明の名称】）
EJ-15・小文字を含む大文字表記を適切に処理していない（【発明の名称】）
EJ-16・クレーム引用に関する表現を適切に翻訳していない（【特許請求の範囲】）

各チェック項目の詳細及び具体例については第 3 章 3.2.1.を参照のこと。

4.3.8. 相対評価方法

「仮設手法」の「人手による品質評価の実施」部分には、前節までで説明した「1. 評価の観点及び基準」の定義に加え、これらの評価観点及び基準に基づいて得られた評価結果の具体的な活用方法として「2. 相対評価方法」と「3. 絶対評価方法」とが定義されている。「特許庁手法」では、これらについても定義内容を再検討した。まず本節にて「2. 相対評価方法」に対する再検討の内容を示す。

4.3.8.1. 「仮設手法」の定義内容

「仮設手法」における「相対評価方法」部分の記載内容は以下のとおりである。

表 4-17. 「仮設手法」「2. 相対評価方法」部分

2. 相対評価方法

1. (1)～(3)の評価結果をそれぞれスコア化し、観点毎に、複数の機械翻訳システムの中で、相対評価を行う。

「特許庁手法」では、上記定義について、「相対評価の評価対象」、「評価結果のスコア化の方法」、「総合評価の導入」という3つの観点から再検討を行なった。

4.3.8.2. 相対評価の対象

「仮設手法」の定義では、相対評価に用いる評価観点として「(1)～(3)」、そして評価対象として「複数の機械翻訳システム」を想定している。まずは、これらの妥当性について再検討した。

〔相対評価に利用する評価結果〕

「仮設手法」では、相対評価に使用する評価結果として「(1) 内容の伝達レベルの評価」「(2) 重要技術用語の翻訳精度の評価」「(3) チェックリストに基づく評価」の全てを想定している（「流暢さの評価」は「仮設手法」では評価観点として存在していない）。

しかし、このうち「チェックリストに基づく評価」については、「特許庁手法」ではこれを相対評価には用いない方針としている（4.3.4.2.）。一方、「特許庁手法」にて新設された「流暢さの評価」は、補助的な指標として相対評価に利用すべきものである。

したがって、定義にはこれらのことを反映する必要があるが、「特許庁手法」では「流暢さの評価」を(3)として挿入し、「内容の伝達レベルの評価」を(4)に繰り下げているため、「(1)～(3)の評価結果」を使用するという記載自体は変更不要である。

[相対評価を行う対象]

「仮設手法」の定義は、相対評価の対象として「複数の機械翻訳システム」のみが想定されているが、「特許庁手法」では、4.1.3.に示したとおり、相対評価の用途として他にも「特定のシステムの翻訳品質の定点観測（同一システムの二つの時点の翻訳精度比較）」等を想定している。したがって、定義内容は以下となる。

表 4-18. 「仮設手法」「2. 相対評価方法」部分

2. 相対評価方法

1. (1)～(3)の評価結果をそれぞれスコア化し、観点毎に、複数の機械翻訳システム間、**もしくは特定のシステムにおける複数の機械翻訳結果間で**、相対評価を行う。

4.3.8.3. 評価結果のスコア化の手順

「仮設手法」では、相対評価のため各観点の評価結果をスコア化することが示されているが、スコア化の具体的な手順は定義されていない。「特許庁手法」では、評価運用時の便宜のため、相対評価に用いられる各評価観点について、評価結果のスコア化の手順を定義することとした。なお、スコアは各観点とも、満点を5に揃える形とした。

【「(1) 内容伝達レベルの評価」のスコア化】

「内容の伝達レベルの評価」は、複数（3名以上）の評価者が、各評価対象文の機械翻訳結果について5段階で評価する形式の評価である。本項目のスコア化において留意すべき点として、評価値を文長別に算出し、これを実際の特許文献の文長比率に基づき補正する必要があることが挙げられる。

「特許庁手法」にて評価対象とする各文は、評価のためにあらかじめ作成された「テストセット」から抽出される。この「テストセット」は、実際の特許文献の文長構成に基づき構成するが（5.2.2.2.参照）、ここから評価対象文を抽出する際は、評価作業の負担軽減のため、難易度の高い長文の比率を最小限とする形の運用も想定される。

このような場合、評価対象文の構成は実際の特許文献に比べて長文の比率が小さいものとなり、その結果をそのまま相対評価に用いると、長文の翻訳に適したシステムが不利を被る。このため、評価結果を文長別に算出し、これを実際の特許文献の構成比率に合わせて補正する必要がある。

上記を踏まえ、「特許庁手法」における本項目のスコア化手順は、以下のとおりとした。

- ① 各文に対し、全評価者の評点の平均値を算出する。
- ② 評価対象全文の平均値を数段階の文長別に算出する。
- ③ 文長別の平均値を、実際の特許文献の文長比率に基づき補正し、これを「内容伝達レベルの評価」のスコアとする。

【「(2) 重要技術用語の翻訳精度の評価」のスコア化】

「重要技術用語の翻訳精度の評価」は、各評価対象文に対して選定された重要技術用語を対象とし、機械翻訳結果における訳語の翻訳品質を A～D で評価するものである。

本評価観点においては、A～D のそれぞれについて、評価対象とした重要技術用語の総語数を母数とした比率を算出し、これらを適訳率、誤訳率等として評価値に用いることとしている（4.3.2.2.）。本項目にて、これらの評価値の算出方法と、そのスコア化の手順について、以下のとおり定義した。

- ① 評価対象とした「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～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」総数
 - ・ 誤訳率 = 評価 C（誤訳）の語数 ÷ 「重要技術用語」総数
 - ・ 不訳率 = 評価 D（不訳）の語数 ÷ 「重要技術用語」総数
- ③ 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 点満点のスコアに換算する。

上記定義では、重要技術用語のスコア換算（③）の具体的な換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。

例えば、可訳語（評点 B）は「粗読」用途・「精読」用途においては有用だが「検索」用途においては有用でない。したがって、それぞれの用途への有用性を評価するためのスコア換算をすれば、「粗読」用途では換算式を例えば「(適訳率 + (可訳率×0.8)) × 5」等として可訳語を評価対象とし、一方「検索」用途では「適訳率 × 5」等として可訳率を排除する（または重み付けを下げる）、といった個別の調整が必要となる。

[[「(3) 流暢さの評価」のスコア化]

「流暢さの評価」は、機械翻訳結果の文としての読みやすさ、理解しやすさを 5 段階で評価するものである。機械翻訳結果の流暢さは原文の文長に影響されることも多いため、スコア化にあたっては、「内容の伝達レベルの評価」と同様、文長構成について配慮する必要がある。

①評価結果を文長別に集計し、これを特許文献の文長比率に基づき補正して、「流暢さの評価」のスコアとする。

[[「(4) チェックリストに基づく評価」について]

「チェックリストに基づく評価」の評価結果は、相対評価及び絶対評価には含めない。このため、スコア化は不要である。

4.3.8.4. 「総合評価」の導入

「仮設手法」では、相対評価は評価観点ごとに行うこととされているが、実際の品質評価においては、各評価観点を総合して最終的な判断を下す必要がある。また、その際は、各評価観点のスコアを同等に扱うのではなく、評価の目的に応じて、それぞれに適切な重み付けを行うべきである。

これを踏まえ、「特許庁手法」では、「仮設手法」の「観点毎に」という定義を削除したうえで、項目末尾に「総合評価」として下記一文を追加した。

総合評価

各評価値を、用途等を考慮して適切に重み付けし、総合評価を行う。

4.3.8.5.「特許庁手法」の確定内容

上記再検討により、「特許庁手法」の「相対評価方法」の定義内容を下表 4-19 のとおり確定した。

表 4-19.「特許庁手法」「2. 相対評価方法」部分

2. 相対評価方法

「1. 評価の観点及び基準」(1)～(3)の各観点の評価結果をそれぞれスコア化し、複数の機械翻訳システム間、もしくは特定のシステムにおける複数の機械翻訳結果間で、相対評価を行う。

※ 「(4) チェックリストに基づく評価」の評価結果は、相対評価には用いない。

各評価結果のスコア化は、以下の手順を基本とする。

(1)「内容伝達レベルの評価」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「内容伝達レベルの評価」の評価値とする。

(2)「重要技術用語の翻訳精度の評価」結果のスコア化

- ①評価対象とした「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～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」総数
 - ・誤訳率 = 評価 C (誤訳) の語数 ÷ 「重要技術用語」総数
 - ・不訳率 = 評価 D (不訳) の語数 ÷ 「重要技術用語」総数
- ③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。

(3)「流暢さの評価」結果のスコア化

- ①評価結果を文長別に集計し、これを特許文献の文長比率に基づき補正して、「流暢さの評価」のスコアとする。

(4) 総合評価

上記(1)～(3)の評価値を、用途等を考慮して適切に重み付けし、総合評価を行う。

4.3.9. 絶対評価方法

「絶対評価」とは、評価対象とした機械翻訳結果の各種特許実務用途への有用性を、品質評価結果から直接判断する方式の評価を指す。具体的な実務用途として、検索用途、粗読用途のほか、実体審査での精読、権利侵害の存否等も想定される。「相対評価」も基本的には特定の用途を想定して行うものであるが、ここで分かるのは、品質評価スコアの比較による評価対象間の優劣のみであり、それぞれの評価対象が各種用途にどの程度有用であるかを把握するためには、スコアと用途とを適切な基準でマッチングさせる必要がある。

「仮設手法」では、この「絶対評価」を行うための手順を「3. 絶対評価方法」として定義している。「特許庁手法」の作成にあたり、この定義について再検討した。以下にその詳細を記す。

4.3.9.1. 「仮設手法」の定義内容

表 4-20 に「仮設手法」における「絶対評価方法」の定義内容を示す。

表 4-20. 「仮設手法」「3. 絶対評価方法」部分

3. 絶対評価方法 (0) 絶対評価スコアと用途とのマッチングは、以下のとおりとする。 スコア 5 : 権利認定、引用発明認定に適した翻訳品質 スコア 4 : 実体審査、先行技術調査における精読に適した翻訳品質 スコア 3 : 先行技術調査及び文献照会時の粗読及び検索データに適した翻訳品質 スコア 2 : 粗読又は検索データに適した翻訳品質 スコア 1 : 特許調査に際しては全く利用できない翻訳品質 (1) 内容の伝達レベルの評価（平均値）に基づいて、仮スコアを定める。 内容伝達 : 4.5 ~ 5.0 → 仮スコア : 5 内容伝達 : 3.5 ~ 4.5 → 仮スコア : 4 内容伝達 : 2.5 ~ 3.5 → 仮スコア : 3 内容伝達 : 1.5 ~ 2.5 → 仮スコア : 2 内容伝達 : 1.0 ~ 1.5 → 仮スコア : 1 (2) 技術用語の翻訳について 技術用語の翻訳に関する評価結果を踏まえ、正解率が所定割合以下の場合は、スコアを 1 ランク下げる。 (3) チェックリストを利用した評価について チェックリストを利用した評価結果に関して、特筆すべき問題点が存在する場合は、スコアを 1 ランク下げる。 また、チェックリストを利用して、翻訳ミスの傾向を分析する。

「仮設手法」の絶対評価方法は、表 4-20 のとおり、機械翻訳結果の具体的な用途に応じて 5 段階の「絶対評価スコア」を定義し、これと「内容の伝達レベル」のスコアとをマッチングさせ、さらに技術用語の翻訳の正解率が一定以下の場合と、「チェックリストに基づく評価」にて特筆すべき問題点が存在する場合に、スコアを 1 ランクずつ下げる、というものである。

「特許庁手法」では、5 段階の「絶対評価スコア」自体は「仮設手法」のものをそのまま採用し、これと品質評価スコアをマッチングさせる手順について再検討した。

4.3.9.2. 「絶対評価スコア」と「内容の伝達レベルの評価」とのマッチング

「仮設手法」では、「絶対評価スコア」とのマッチングに「内容の伝達レベルの評価」の評価スコアのみを使用している。

「内容の伝達レベルの評価」は、特許文献の機械翻訳結果において最も重要な「内容の正確さ」の観点からの評価であり、また複数の評価者による重複評価であって信頼度も高いことから、これを「絶対評価スコア」とのマッチング対象とすることは妥当と考える。

なお、「仮設手法」には「絶対評価スコア」と「内容の伝達レベルの評価」のマッチング基準が（1）に示されているが、この妥当性については、第 8 章にて、「特許庁手法」の実践結果を踏まえて検証することとし、当面は「仮設手法」のマッチング基準をそのまま採用することとする。

4.3.9.3. 「重要技術用語の翻訳精度の評価」の反映方法

「絶対評価スコア」に列挙された各種用途のうち、「検索」用途のみは、たとえ内容の伝達レベルが低くても、重要な技術用語さえ適訳に訳されていれば、一定の有用性を得られる（逆に言えば、重要技術用語の翻訳品質が一定以下であれば、いくら内容の伝達レベルが高くても検索用途への有用性は低い）という、他の用途と異なる性質をもつ。

「仮設手法」における重要技術用語の翻訳精度の評価スコアの扱いは、これが一定値以下であった場合に、「内容の伝達レベルの評価」に基づき定まった絶対評価スコアを 1 ランク下げるというものである。だが、「絶対評価スコア」においては、検索用途への有用性がなければスコアが 3 以上となることはなく、あればスコア 2 が確実に保証されるわけであるから、これをそのまま反映し、「重要技術用語の翻訳精度の評価」の「適訳率」が一定値未満である場合はスコア 2 を上限とし、一定値以上であればこれを下限とする、という形で反映させるほうがより合理的である。

したがって「特許庁手法」では、「重要技術用語の翻訳精度の評価」の反映手順に関する記載内容を以下のとおり改めた。

(2)「重要技術用語の翻訳精度の評価」の「適訳率」が一定値未満の場合は、絶対評価スコアの上限を 2 とし、これを超えている場合はスコアを 2 に下げる。一方、「適訳率」が一定値以上の場合は、絶対評価スコアの下限を 2 とし、これに満たない場合はスコアを 2 に上げる。

なお、「適訳率」の一定値をどこに定めるかについては、第 8 章にて、「特許庁手法」の実践結果を踏まえて考察する。

4.3.9.4.「チェックリストに基づく評価」及び「流暢さの評価」の取り扱い

「仮設手法」では、「チェックリストに基づく評価」に関しても、評価結果に「特筆すべき問題点が存在する場合はスコアを 1 ランク下げる」こととしている。

しかしながら、「特許庁手法」では、この評価観点はフィードバック利用のためのものと位置づけ、相対評価・絶対評価には用いない方針としており(4.3.4.2.参照)、この方針に基づいて評価基準等も設定している。このため、「仮設手法」の上記定義は廃止し、「チェックリストに基づく評価」は絶対評価の対象外とした。

一方、「流暢さの評価」は「特許庁手法」にて新設された評価観点であるため、「仮設手法」では当然ながら言及されていない。

「流暢さの評価」は、相対評価において「内容の正確さ」が同等である場合の優劣判断に用いる等の補助的な指標であり、「絶対評価」に加点要素として組み込む必要はない。しかしながら、評価自体は「内容の伝達レベルの評価」及び「重要技術用語の翻訳精度の評価」と同時に実施される想定であり、そこで特筆すべき問題点が検出された場合に適宜減点することに問題はない。このため「特許庁手法」には、これを許容する文言を追加した。

(3)「流暢さの評価」において、特筆すべき問題点が存在する場合は、その内容に応じてスコアを減点する。

4.3.9.5. 「特許庁手法」の暫定内容

以上の検討により、「特許庁手法」における「絶対評価方法」の定義は下表 4-21 のとおりとした。ただし、4.3.9.2.で述べたとおり、内容伝達スコアと絶対評価スコアとのマッチング基準は「仮設手法」の内容をそのまま採用しており、その妥当性は「特許庁手法」の実践結果を踏まえて検証する予定であるため、ここでの定義は暫定的なものである。

表 4-21. 「特許庁手法」「3. 絶対評価方法」部分 ※暫定

3. 絶対評価方法

(0) 絶対評価スコアと用途とのマッチングは、以下のとおりとする。

スコア 5：権利認定、引用発明認定に適した翻訳品質

スコア 4：実体審査、先行技術調査における精読に適した翻訳品質

スコア 3：先行技術調査及び文献照会時の粗読及び検索データに適した翻訳品質

スコア 2：粗読又は検索データに適した翻訳品質

スコア 1：特許調査に際しては全く利用できない翻訳品質

(1) 「内容の伝達レベルの評価」の**スコア**に基づき、絶対評価スコアを定める。

内容伝達スコア：4.5～5.0 → **絶対評価スコア：5**

内容伝達スコア：3.5～4.5 → **絶対評価スコア：4**

内容伝達スコア：2.5～3.5 → **絶対評価スコア：3**

内容伝達スコア：1.5～2.5 → **絶対評価スコア：2**

内容伝達スコア：1.0～1.5 → **絶対評価スコア：1**

(2) 「重要技術用語の翻訳精度の評価」の「**適訳率**」が一定値未満の場合は、絶対評価スコアの上限を**2**とし、これを超えている場合はスコアを**2**に下げる。一方、「**適訳率**」が一定値以上の場合は、絶対評価スコアの下限を**2**とし、これに満たない場合はスコアを**2**に上げる。

(3) 「**流暢さの評価**」において、特筆すべき問題点が存在する場合は、その内容に応じてスコアを減点する。

4.3.10. 主要項目と補助項目

「特許庁手法」では、「1. 評価の観点及び基準」にて4つの評価観点を定義した。評価を実施する際は、その全てを実施することにより、多角的な観点から機械翻訳の品質を評価することが可能となる。しかしながら、評価作業のリソース（評価者の人数、時間等）の制限等により、全ての評価項目を実施することが困難な場合も考えられる。

「特許庁手法」では、このような場合の対処策として、4つの評価項目を「主要項目」と「補助項目」とに区分し、全ての評価項目の実施が困難な場合は「主要項目」のみを実施可能とすること、及び一部評価項目については、実施条件を緩めることを定め、これを新規項目として追加した。

4.3.10.1 主要項目と補助項目の区分

「特許庁手法」の各評価観点を、以下の理由により主要項目と補助項目に区分した。

「(1) 内容の伝達レベルの評価」…【主要項目】

本評価項目は、機械翻訳結果の「内容の正確さ」を評価するものであり、「特許庁手法」においてはメイン評価といえる。相対評価、絶対評価時のいずれにも必須の評価観点であり、【主要項目】として扱う。

「(2) 重要技術用語の翻訳精度の評価」…【主要項目】

「特許庁手法」の想定する機械翻訳結果の主要な用途のひとつである「検索」用途においては主要な評価観点となる。このため相対評価のみならず絶対評価においても必須の観点であり、同じく【主要項目】として扱う。

「(3) 流暢さの評価」…【補助項目】

「特許庁手法」においては補助的な評価観点であり、相対評価においては上記2つの評価観点で優劣がつかない場合に参照するための観点、絶対評価においても参考情報の扱いであり、リソースに制限がある場合には省略しても影響は小さい。

「(4) チェックリストに基づく評価」…【補助項目】

他の項目と異なり、主にフィードバック利用のために設定された評価観点であり、相対評価や絶対評価には原則として関与しない。評価の難易度、そして評価者に求められる能力ともに他の評価観点に比べて高く、また作業負担も格段に大きい。したがって、品質評価の目的が相対評価や絶対評価である場合は、本評価を実施する必要があるか慎重に検討すべきである。

4.3.10.2「内容の伝達レベルの評価」の実施要件の緩和

「内容の伝達レベルの評価」は、各文について 3 名以上の評価者による重複評価が要件となっている。また、各文について、その文の属する技術分野に通じた者を評価者に起用することも求められている。通常、特許文献の機械翻訳結果の有用性評価は全技術分野が対象となることを考えると、1 回の評価に必要な評価者の人数は多人数となり、確保が困難なケースも想定される。

「内容の伝達レベルの評価」は「特許庁手法」の主要な評価項目であり、重複評価によりその客観性、信頼度を高めることが望ましいことは言うまでもないが、評価の客観性や信頼度は、十分な評価対象文数、評価基準の工夫、客観性の高い「重要技術用語の翻訳精度の評価」の併用等によっても配慮しており、1 文あたり 1～2 名による評価であっても大きく精度を損なうものではない。これを踏まえ、「特許庁手法」では、必要に応じて要件を緩和することを許容する文言を本項目に記載することとする。

4.3.10.3「特許庁手法」の確定内容

上記検討を踏まえて「特許庁手法」に追加した「主要項目と補助項目」の項目内容を以下に示す。

表 4-22. 「特許庁手法」「4. 主要項目と補助項目」部分

4. 主要項目と補助項目

前記 1. (1)～(4)の各評価項目を、以下のとおり「主要項目」と「補助項目」とに区分し、評価作業のリソース（人数、時間等）に制限があり、全ての評価項目を実施することが困難な場合には、「主要項目」のみを実施するものとする。

主要項目：(1) 内容の伝達レベルの評価
(2) 重要技術用語の翻訳精度の評価

補助項目：(3) 流暢さの評価
(4) チェックリストに基づく評価

また、主要項目である「(1) 内容の伝達レベルの評価」において、1 文に対し 3 名以上の評価者を用意することが不可能な場合は、評価者の人数を削減してこれを実施することも可とする。

4.4. 特許庁機械翻訳品質評価手法の内容（２）：「事前準備」部分

「特許庁手法」は、大きく「事前準備」部分と「人手による品質評価の実施」部分とに区分される。このうち「事前準備」部分は、評価実施のためのテストセットの要件等を定義するものであり、その内容は、評価の観点や基準といった評価手法本体を定義する「人手による品質評価の実施」部分の内容に応じて定める必要がある。

本項では、前項にて確定した「人手による品質評価の実施」部分の内容を踏まえて再検討した「事前準備」部分の定義内容について説明する。

4.4.1. テストセットの準備

「仮設手法」の「事前準備」部分では、「テストセットの準備」「評価作業用データの準備」「人手評価補助ツールの準備」の３点が定められている。本節では、このうち「テストセットの準備」部分の再検討結果を示す。

4.4.1.1. 「仮設手法」における定義内容

「仮設手法」の「事前準備」部分において、「テストセットの準備」に相当する記載は、下表 4-23 に示すとおりである。

表 4-23. 「仮設手法」の「テストセットの準備」部分 定義内容

１．パテントファミリー関係等を利用して、技術的内容が対応する文献対から技術的内容が対応する文対を抽出することにより、可能な限り、技術分野（IPC）、文献種別、原文の長さ、記載箇所、出願時期、出願人等に偏りなく、機械翻訳対象文（翻訳元の言語）と参照用翻訳文（翻訳先の言語）とのセンテンス単位のペアを作成し、それを集積したテストセットを用意する（言語対ごとに 1000 文程度）。

※機械翻訳対象文に問題（助詞の欠落や誤字・脱字等）があるものは、機械翻訳の適切な評価ができないため、テストセットからは除外する。

※パテントファミリー関係の文献であっても、常に対応する文章となっていないため、テストセットの用意にあたっては、機械翻訳対象文と参照用翻訳文との対応関係についての人手によるチェックを実施する。

※テストセット作成時に、特許文献機械翻訳において重要と想定される技術用語及びチェックリストに該当する対象文を特定しておく。

4.4.1.2. パテントファミリーの利用

上表 4-23 に示したとおり、「仮設手法」では、テストセットはパテントファミリー（対応特許）関係等の既存の特許文献対訳文を利用して作成するものと定められている。

この手法には、既存の対訳文を用いることでテストセットの作成コストを低く抑えられるという大きなメリットがある。しかしその反面、こうした既存の対訳文は、評価対象の機械翻訳システムのチューニング、特に統計翻訳方式（SMT）のシステムの学習データにも利用されている可能性があり、その場合、評価の条件が公平でなくなり正確な品質評価とならない。

したがって、理想的には、テストセットには対訳文の存在しない特許文を採用し、参照用の対訳文は人手翻訳で新規に作成すべきである。しかし、品質評価に費やせるコストには限度があることが予想されるため、少なからぬ作成コストを要するテストセットを必須要件とするのも現実的ではない。また、学習データを用いないタイプのシステムのみを評価する場合や、SMT システムであっても学習データが評価対象文を含まないことが確実である場合であれば、テストセットはパテントファミリーを利用したもので充分である。

「特許庁手法」では、これらの事情を踏まえ、テストセットの作成にあたっては、“対訳文の存在しない特許文を採用し、対訳文は新規に人手翻訳で作成する”ことを理想としつつも、これが困難な場合はパテントファミリー等の既存の対訳文を使用することも許容した。

1. 実際の特許文献から機械翻訳対象文（翻訳元の言語）を選定し、対応する参照用翻訳文（翻訳先の言語）とのセンテンス単位のペアを作成し、評価用テストセットを用意する（言語対ごとに 1,000 文程度）。…

：

：

- ※ 機械翻訳対象文は、対応する言語による対応特許の存在しない特許文献から選定し、人手翻訳にて参照用翻訳文を新規作成することが理想的であるが、テストセットの用意に使用できるリソースに制限がある場合等は、パテントファミリー関係を利用して抽出した対訳文対を用いてもよい。

4.4.1.3. パテントファミリー対訳文の採用基準

「仮設手法」では、パテントファミリー対訳文の採用基準として、“原文が正確であり、かつ対訳文が原文の内容と正しく対応していること”を要件としている。「特許庁手法」も、前述のとおりパテントファミリーをテストセットに使用することを許容しており、この要件定義についても検討が必要である。

「仮設手法」の要件は、内容的には至極妥当なものである。原文が不正確では機械翻訳システムの性能を正しく評価することは不可能であり、対訳文が不正確では、これを参照して行われる評価結果の信頼性も低下する。

しかしながら、実際のパテントファミリー対訳には、原文で省略されている語の補足や、原文に存在する語の省略、構成要素への付番の追加や変更などの微調整が頻繁に行われており、完璧に対応している文は意外に少ない。このため、オリジナルの状態で正しく対応している文のみを選定するとなると、テストセット作成の効率が非常に悪くなる。

「特許庁手法」は、実際の特許文を扱う機械翻訳システムを評価するためのものであるため、テストセットに含める原文も、実際の特許文をそのまま用いるべきである。このため、原文を修正してテストセットに採用することは、どのような形であれ推奨されない。しかし、対訳文に関しては、評価の際の参照のためのものであって機械翻訳の対象ではないため、オリジナルの内容にこだわる必然性は薄い。軽微な修正によって正しい対応が得られるのであれば、そうするほうが効率的である。

この考えに基づき、「特許庁手法」では、原文の修正を禁じる一方で、対訳文に関しては、修正を施したうえでこれを採用することを許容することとした。

※ パテントファミリー関係の文献であっても、常に対応する文章となっているとは限らないため、テストセットの用意にあたっては、機械翻訳対象文と参照用翻訳文との対応関係についての人手によるチェックを実施し、対応関係に問題がある場合はこれを除外する。ただし、**参照用翻訳文を修正することで正しい対応が得られる場合は、修正を施したうえで採用してもよい。なお、機械翻訳対象文の修正は認めない。**

4.4.1.4. 重要技術用語の選定および該当チェック項目の特定

「特許庁手法」では、評価観点「(2) 重要技術用語の翻訳精度の評価」の検討にあたり、評価対象とする重要技術用語は、テストセットの各文に対して最大 3 語、あらかじめ選定されている前提とした (4.3.2.4.)。

また、「(4) チェックリストに基づく評価」では、機械翻訳における言語共通の課題や、翻訳言語に応じた課題についてもチェック項目として多数設定しているが (4.3.5.~7.)、テストセットの各文がそれぞれのチェック項目の検証に適したものであるかがあらかじめ特定されていれば、テストセットから評価対象文を抽出する際に、各チェック項目の検証に必要な文を選定することも容易になる。

「特許庁手法」では、これらを踏まえ、テストセットの各文について重要技術用語をあらかじめ選定してこれを記憶しておくことと、各チェックリスト項目の検証に適した文をあらかじめ特定しておくことを、テストセットの準備の際の要件として追加した。

※ テストセット作成時に、各文に含まれる重要技術用語をあらかじめ選定し、評価作業時に提示するためにデータ上に記憶しておく。さらに、チェックリストの【任意項目】や【翻訳言語別任意項目】等の検証に適した文についても、あらかじめこれを特定し、各項目に該当する文がテストセットに一定数含まれていることを確認する。

4.4.1.5. 「特許庁手法」の確定内容

以下、「特許庁手法」の「テストセットの準備」部分の定義内容を示す。

表 4-24. 「特許庁手法」の「テストセットの準備」部分 定義内容

1. 実際の特許文献から機械翻訳対象文（翻訳元の言語）を選定し、対応する参照用翻訳文（翻訳先の言語）とのセンテンス単位のペアを作成し、評価用テストセットを用意する（言語対ごとに 1,000 文程度）。テストセットに選定する文は、可能な限り、技術分野（IPC）、文献種別、原文の長さ、記載箇所、出願時期、出願人等に偏りのないようにする。
- ※ 機械翻訳対象文は、対応する言語による対応特許の存在しない特許文献から選定し、人手翻訳にて参照用翻訳文を新規作成することが理想的であるが、テストセットの用意に使用できるリソースに制限がある場合は、パテントファミリー関係を利用して抽出した対訳文対を用いてもよい。
 - ※ 機械翻訳対象文に問題（誤字・脱字や、複数の解釈が可能な曖昧な係り受け等）があるものは、機械翻訳の適切な評価ができないため、テストセットからは除外する。
 - ※ パテントファミリー関係の文献であっても、常に対応する文章となっているとは限らないため、テストセットの用意にあたっては、機械翻訳対象文と参照用翻訳文との対応関係についての人手によるチェックを実施し、対応関係に問題がある場合はこれを除外する。ただし、参照用翻訳文を修正することで容易に対応可能な場合は、修正を施したうえで採用してもよい。なお、機械翻訳対象文の修正は認めない。
 - ※ テストセット作成時に、各文に含まれる重要技術用語をあらかじめ選定し、評価作業時に提示するためにデータ上に記憶しておく。さらに、チェックリストの【任意項目】や【翻訳言語別任意項目】等に該当する文についても、あらかじめこれを特定し、各項目に該当する文がテストセットに一定数含まれていることを確認する。

4.4.2. 評価作業用データの準備

品質評価は、テストセット全件に対して行うのではなく、そこからピックアップされた 200 文程度の機械翻訳対象文について、これらを品質評価対象の機械翻訳システムによって機械翻訳した結果、すなわち「評価作業用データ」に対して行う想定である。「事前準備」部分は、この「評価作業用データの準備」についても定義している。

4.4.2.1. 「仮設手法」の定義内容

「仮設手法」では「評価作業用データの準備」について以下のとおり定めている。

表 4-25. 「仮設手法」の「評価作業用データの準備」部分 定義内容

- | |
|--|
| <p>2. テストセットから評価対象件数の機械翻訳対象文を抽出する（評価対象件数は 200 文程度）。機械翻訳対象文の抽出は、テストセットの作成と同様に、技術分野等について偏りないように行うこと。</p> <p>3. 機械翻訳対象文について、品質評価の対象となる機械翻訳システムによって、機械翻訳を実施する。</p> |
|--|

4.4.2.2. 「評価作業用データ」の構成

「仮設手法」では、評価作業用データの構成は、評価対象文数は 200 文程度、かつ、技術分野等について偏りなく構成するよう定めている。

しかしながら、品質評価に適したデータ構成は、評価の目的や実施する評価観点によって異なり、必ずしも一定ではない。例えば、フィードバックのための「チェックリストに基づく評価（4.3.4.～7.）」を行う場合、1 文あたりの作業負担が大きいことを考慮してデータ件数を減らしたり、分析の難易度の高い長文の比率を下げる等の調整を行うべきであるし、性能の極めて接近したシステム同士の相対評価を行う場合等、200 文では有意な相対順位が得られないことが予測される場合もある。

このため「特許庁手法」では、評価作業用データの構成について、「仮設手法」の定義に準拠しつつも、これを評価の目的等に応じて調整可能とすることとした。

4.4.2.3. 「特許庁手法」の確定内容

上記を反映した「特許庁手法」の定義内容を下表 4-26 に示す。

表 4-26. 「特許庁手法」の「評価作業用データの準備」部分 定義内容

- | |
|--|
| <p>2. テストセットから評価対象件数の機械翻訳対象文を抽出する（評価対象件数は 200 文程度）。機械翻訳対象文の抽出は、テストセットの作成と同様、技術分野等について偏りないように行うことを原則とする。ただし、評価の目的や評価作業のリソースに配慮した調整は適宜行ってよい。</p> <p>3. 機械翻訳対象文について、品質評価の対象となる機械翻訳システムによって、機械翻訳を実施する。</p> |
|--|

4.4.3. 評価支援ツールの準備

「仮設手法」の「事前準備」部分では、評価作業に用いる人手評価補助ツールを用意することについても定められている。

表 4-27. 「仮設手法」の「評価支援ツールの準備」部分 定義内容

- | |
|--|
| <p>4. 機械翻訳結果の評価に際し、評価を簡便に実施するために必要なツール（人手評価補助ツール）の作成をする。本ツールは、機械翻訳対象文、参照用人手翻訳、及び機械翻訳対象文の機械翻訳結果からなる品質評価対象データをインポートし、評価者にシート形式で提示するとともに、以下のような機能を有する。</p> <ul style="list-style-type: none">・機械翻訳結果に 5 段階評価を行う場合、ワンクリック等の簡潔な操作で任意の段階の評価を行うことができる。・機械翻訳対象文や機械翻訳結果について、評価者が、備忘録や評価根拠のコメントを付すことができる。・機械翻訳対象文について、評価者の指定範囲を自動的に辞書引きすることができる。 |
|--|

この定義内容に関しては、特に変更すべき点は見当たらないため、そのまま「特許庁手法」に採用した。

なお、本調査にて作成した評価支援ツールの詳細は、7.1.4.及び 6.3.3.を参照のこと。

第 5 章

英日機械翻訳に関するテストセットの作成

第5章 英日機械翻訳に関するテストセットの作成

本調査では、「特許庁手法」にて定めたテストセットの定義に基づき、英日機械翻訳に関するテストセットを作成した。以下にその内容を記す。

5.1. テストセットの要件

本調査にて作成した英日機械翻訳に関するテストセット（以下、「英日テストセット」という。）の要件について本項にまとめた。

5.1.1. 「特許庁手法」のテストセット定義

英日テストセットは、「特許庁手法」に定義されたテストセットの要件（4.4.1.5.）に従い作成した。「特許庁手法」のテストセットに関する記載は以下のとおりである。

1. 実際の特許文献から機械翻訳対象文（翻訳元の言語）を選定し、対応する参照用翻訳文（翻訳先の言語）とのセンテンス単位のペアを作成し、評価用テストセットを用意する（言語対ごとに 1,000 文程度）。テストセットに選定する文は、可能な限り、技術分野（IPC）、文献種別、原文の長さ、記載箇所、出願時期、出願人等に偏りのないようにする。

※ 機械翻訳対象文は、対応する言語による対応特許の存在しない特許文献から選定し、人手翻訳にて参照用翻訳文を新規作成することが理想的であるが、テストセットの用意に使用できるリソースに制限がある場合等は、パテントファミリー関係を利用して抽出した対訳文対を用いてもよい。

※ 機械翻訳対象文に問題（誤字・脱字や、複数の解釈が可能な曖昧な係り受け等）があるものは、機械翻訳の適切な評価ができないため、テストセットからは除外する。

※ パテントファミリー関係の文献であっても、常に対応する文章となっているとは限らないため、テストセットの用意にあたっては、機械翻訳対象文と参照用翻訳文との対応関係についての人手によるチェックを実施し、対応関係に問題がある場合はこれを除外する。参照用翻訳文を修正することで正しい対応が得られる場合は、修正を施したうえで採用してもよい。なお、機械翻訳対象文の修正は認めない。

※ テストセット作成時に、各文に含まれる重要技術用語をあらかじめ選定し、評価作業時に提示するためにデータ上に記憶しておく。さらに、チェックリストの「特殊ケース」や「翻訳言語に応じた項目」等に該当する文についても、あらかじめこれを特定し、各項目に該当する文がテストセットに一定数含まれていることを確認する。

5.1.2. 英日テストセットの要件

英日テストセットは、上記「特許庁手法」の定義に従い、1,000 文にて構成することとした。また、「チェックリストに基づく評価」における【任意項目】(4.3.6.) や、英日機械翻訳における特有課題及び留意点に対応した【翻訳言語別任意項目 [英日]】(4.3.7.) について検証可能な文が十分な量含まれるよう配慮した。

なお、「特許庁手法」では、一つめの注記にあるとおり、テストセットに含める機械翻訳対象文は「対応特許の存在しない特許文献から選定し、人手翻訳にて参照用翻訳文を新規作成する」ことを理想としている。しかしながら、本調査では、調査期間等の制約により、参照用翻訳文を人手翻訳で準備することが困難であり、上記の理想的な条件で英日テストセットを作成することが不可能であった。このため、同注記にてリソースに制限がある場合等に許容されている“パテントファミリー関係を利用して抽出した対訳文対を用いる”手法にてこれを作成することとした。

それに伴い、三つめの注記事項、すなわち「パテントファミリーから選定した文対であっても、常に対応する文章となっているとは限らない」ことに対する配慮が必要となるため、同注記に定められているとおり、機械翻訳対象文の選定時に対訳文の対応関係をチェックし、問題がある場合はこれを除外、もしくは参照用翻訳文の修正を行うこととした。

5.2. 英日テストセットの作成手順

英日テストセットは、パテントファミリー関係を利用し、以下の手順にて作成した。

- ① 英日文アライメントの作成と英日テストセット候補データの抽出
- ② 英日テストセットにおける文の比率の確定
- ③ 文の選定と目視チェック
- ④ チェックリスト該当項目の記録
- ⑤ 重要技術用語の特定

本項では、上記各工程について詳述する。

5.2.1. 英日文アライメントデータの作成と英日テストセット候補データの抽出

英日文アライメントデータは、米日、欧日別に EPO の書誌データベース DOCDB のパテントファミリー情報で対応する文献を特定、その明細書フルテキストを取得し、NICT の文アライメントツール¹を用いて作成した²。作成した英日文アライメントデータのうち、対応度を示すスコアが一定値以上（0.08）のものを 1 文献あたりランダムに 15 文程度抽出し、これを「英日テストセット候補データ」とした。

なお、テストセットに含めるべき文例が英日テストセット候補データ中に存在しない場合に備え、これを効率的に検索・抽出できるよう、英日文アライメントデータをデータベース化し、データ不足時は適宜ここからデータを検索して候補データに追加した。

5.2.2. 英日テストセットにおける文の比率の確定

「特許庁手法」では、テストセットに含める文は「可能な限り、技術分野（IPC）、文献種別、原文の長さ、記載箇所、出願時期、出願人等」に偏りのないよう選定することが定められている。なお、ここで言う「偏りのない」とは、実際の特許文献の状況に則した配分を意味しており、均等配分を意味するものではない。

本調査では、実際の特許文献の状況に則した配分を行ううえで、「記載箇所」、「原文の長さ」「技術分野（IPC）」「文献種別」の順に比率を決定し、テストセットの構成を細分化した。一方、「出願時期」「出願人」については、1 文献からの選定文数を原則 1～2 文程度³にとどめることで対応した。

¹ このツールは英日対訳辞書に基づいて日本語文章ファイルと英語文章ファイルから対訳関係にある文を抽出する。抽出した英日文アライメントデータには英日対訳辞書に基づく文の対応度（スコア）が付けられる。

² 米日については同様の方法で作成した既存の文アライメントを利用した。

³ 特定の技術分野において必要数のパテントファミリー数が得られなかった場合等、3 文以上を選定している場合もある。

以下、英日テストセットに適用した項目別の文の比率について示す。

5.2.2.1. 記載箇所

英語特許文献は、その記載箇所、すなわち「発明の名称（タイトル）」「要約」「特許請求の範囲（クレーム）」「発明の詳細な説明（明細書）」のそれぞれにおいて、特徴のある文型や特有表現が見られる。このため英日テストセットも、各記載箇所から充分かつ適切な量の文を採用する必要がある。

英日テストセットにおける記載箇所別の構成比率を定めるにあたり、まずは第 3 章「英日機械翻訳の現状分析」にて使用したサンプル文献 20 文献における構成比率を調査した。表 5-1 にその結果を示す。

表 5-1. サンプル 20 文献の記載箇所別構成比

	タイトル	要約	クレーム	明細書	合計
文 数	20 文	64 文	295 文	1,615 文	1,994 文
構成比	1%	3%	15%	81%	100%

サンプル文献 1（US2007000673）における構成比率は 1:1:21:86 文であり、上表の数値はおおむね妥当と考えられる。ただし、この構成比率をそのまま英日テストセットに採用すると、「タイトル」や「要約」からの採用文は、1,000 文のテストセットにおいてはそれぞれ 10 文、30 文と非常に少なくなり、十分なバリエーションが確保できない。また、特許文献には、例えば中国の CPA や韓国の KPA など、公式な英語データは「タイトル」と「要約」のみ存在するというケースも多いため、これらを考慮して、英日テストセットにおいては、「タイトル」と「要約」の比率を、上記調査結果よりも高く設定することとした。さらに「クレーム」についても、特許実務における重要性を考慮し、こちらも実際の構成比よりも若干高めとした。

以上の判断により、英日テストセットの記載箇所別構成比率は、下表 5-2 のとおり定めた。

表 5-2. 英日テストセット記載箇所別構成比

	タイトル	要約	クレーム	明細書	合計
文 数	50 文	50 文	200 文	700 文	1,000 文
構成比	5%	5%	20%	70%	100%

5.2.2.2. 原文の長さ

英日テストセットの文長別比率についても、実際の特許文献における構成比に従うべく、前出のサンプル 20 文献における文長別構成比率を調査した。なお、英文では、文構造の複雑さは文字数よりも語数に比例するため、文長構成は語数を基準とし、51 語以上の文を「長文」、26 語～50 語の文を「中文」、25 語以下の文を「短文」と定めた。表 5-3 に、サンプル 20 文献の文長別構成比率を示す。

表 5-3. サンプル 20 文献 文長（語数）別構成比

	タイトル	要約	クレーム	明細書	合計	
1～25 語	20 文	32 文	512 文	892 文	1,096 文	55%
26～50 語	0 文	25 文	99 文	617 文	741 文	37%
51 語～	0	7 文	44 文	106 文	157 文	8%

なお、上記集計値は、データ処理の都合上、文中に「:」や「;」が含まれる場合、そこで文を区切り、複数の文としてカウントしている。このため、実際の構成比よりもやや短文寄りに偏っていると考えられる（ただし、多くの機械翻訳エンジンはこれと同様の文区切り基準を用いている）。

英日テストセットでは、この事情を踏まえて短文の比率を若干下げ、代わりに中文、長文の比率を上げる調整を行なった。下表 5-4 に、英日テストセットにて採用した語数別構成比を示す。

表 5-4. 英日テストセット 語数別構成比

	短文（1～25 語）	中文（25～50 語）	長文（51 語～）	合計
文 数	500 文	400 文	100 文	1,000 文
構成比	50%	40%	10%	100%

長文 100 文に関しては、「記載箇所」別の構成にも配慮し、実際の特許文献における数値を参考に、「要約」「クレーム」「明細書」からそれぞれ 1:3:6 程度の比率で選定した。短文と中文は、サンプル数が十分に多いため、「記載箇所」別の配分は省略した。

5.2.2.3. 技術分野（IPC）

技術分野（IPC）に関しては、セクション（A～H）レベルで、実際の英語特許文献における構成比を調査し、それを参考に配分した。

「記載箇所」や「原文の長さ」の根拠としたサンプル 20 文献は、技術分野に関しては恣意的に各セクション均等に選定しており、ここでは使用できない。このため、英語特許文献の技術分野別分布に関しては、米国及び欧州の近年の公開公報（2008 年～2012 年）における IPC セクション別文献数を、民間の特許文献データベースを利用して調査した結果を根拠に用いた。調査結果を下表 5-5 に示す。

なお、カウントは文献に付与された全ての IPC を対象としているため⁴、合計件数は文献総数よりも多くなる。

表 5-5. 米国・欧州公開公報（2008 年～2012 年）IPC セクション別構成比

IPC セクション	US 公開	EP 公開	合計	
A	302,540	204,141	506,681	16.0%
B	286,171	184,008	470,179	14.9%
C	221,613	174,760	396,373	12.5%
D	11,956	15,560	27,516	0.9%
E	40,268	32,148	72,416	2.3%
F	131,849	103,956	235,805	7.5%
G	551,040	217,973	769,013	24.3%
H	452,733	230,542	683,275	21.6%
合計	1,998,170	1,163,088	3,161,258	100.0%

英日テストセットの技術分野別構成比は、上記比率を参考とした。ただし D セクション、E セクションについては、比率に忠実に配分すると文数が 1,000 文中 9 文、23 文と極めて少量となり、データ量が不足する懸念がある。

このため、英日テストセットにおいては、最低限、評価対象データ 200 文を 8 セクション均等文数で構成することが可能なよう、各セクションの文数の下限を 25 文とし、それを超える分について、上記構成比に従い配分することとした。配分結果は表 5-6 のとおりである。

表 5-6. 英日テストセット IPC セクション別構成比

IPC セクション	A	B	C	D	E	F	G	H	合計
文 数	157 文	146 文	123 文	25 文	25 文	74 文	238 文	212 文	1,000 文
構成比	15.7%	14.6%	12.3%	2.5%	2.5%	7.4%	23.8%	21.2%	100%

⁴ 1 文献に同じセクションから複数のコードが付与された場合は 1 カウントとなる。

この構成比を、5.2.2.1.で定めた記載箇所別（タイトル 5%、要約 5%、クレーム 20%、明細書 70%）に配分すると以下となる。

表 5-7. 英日テストセット IPC セクション別構成比

IPC セクション	A	B	C	D	E	F	G	H	合計
文 数	157 文	146 文	123 文	25 文	25 文	74 文	238 文	212 文	1,000 文
タイトル	8 文	7 文	6 文	1 文	1 文	4 文	12 文	11 文	50 文
要 約	8 文	7 文	6 文	1 文	1 文	4 文	12 文	11 文	50 文
クレーム	31 文	29 文	25 文	5 文	5 文	15 文	48 文	42 文	200 文
明細書	110 文	103 文	86 文	18 文	18 文	51 文	166 文	148 文	700 文
構成比	15.7%	14.6%	12.3%	2.5%	2.5%	7.4%	23.8%	21.2%	100%

英日テストセットは、上記構成比にて作成した。理想的には、各セクション内のクラスやサブクラスの比率にも配慮すべきであるが、英日テストセット用の文の選定は、前述の「記載箇所」や「原文の長さ」、さらには「チェックリストに基づく評価」における【任意項目】や【翻訳言語別任意項目 [英日]】の検証に適した文を一定数含める等、多数の考慮事項があり、IPC クラスやサブクラスを限定しつつこれらの条件を満たす文を選定することは作業上困難と判断した。

5.2.3.4. 文献種別

「文献種別」に関しては、2013 年 2 月時点の DOCDB バックファイルのファミリーID 情報に基づき、米日対応文献（345,367 件）と、欧日対応文献で、かつ米日対応文献に該当しないもの（46,016 件）をカウントし、両者の比率に従い、米日対応文献からの文を 883 文（88.3%）、欧日文献からの文を 117 文（11.7%）とした。

表 5-8 文献種別（米国／欧州）別構成比

	DOCDB		テストセット文数
米日対応	345,367 件	88.3%	883 文
欧日対応（米日除く）	46,016 件	11.7%	117 文
合 計	391,383 件	100.0%	1,000 文

なお、欧州文献については構成比率が少ないこともあり、117 文全てを明細書から採取した。

5.2.3. 文の選定と目視チェック

5.2.2.で定めた構成比率に従い、英日テストセット候補データからテストセット用の文を選定した。選定作業は文献単位で行い、候補データ中で最もテストセットに適すると思われる文を1文献あたり1～2文ずつ選定した。選定にあたっては、以下の必須条件と優先条件を満たしているかを目視チェックした。

- ① 原文の内容が文法的に正しい【必須】
- ② 原文と翻訳文の対応が正確である【必須】
- ③ 特許文献の用途上有用である【必須】
- ④ チェックリスト任意項目の検証に適している【優先】

以下、上記の各必須条件、優先条件の詳細について説明する。

5.2.3.1. 原文の内容が文法的に正しい【必須】

原文にスペルミスがある文や、文法的に不正確、またはその疑いがある文は、テストセット用途には向かないため、選定対象から除外した。以下に、除外した文の一例を示す。

原 文	The proximity detector is operable to detect when a finger of a user is proximate the actuator and initiate a second function of the electronic apparatus.
翻訳文	接近検出器は、ユーザの指が作動器に 接近する とそれを検出して、電子機器の第2の機能を始動させる。

proximate が目的語をとる際は to を伴うのが一般的であり、上例の原文は文法的に不正確である疑いがある。このため、英日テストセットの選定対象から除外した。

5.2.3.2. 原文と翻訳文の対応が正確である【必須】

原文に対応する翻訳文は、評価作業時には原文の内容を示すリファレンスとして機械翻訳結果と直接比較されるものであるため、原文の内容と意味的に正確に対応している必要がある。このため、対応の不正確なものは選定対象から除外した。

なお、文中の一部の対応のみが不正確であり、翻訳文の修正により容易に正確な対応に改善できるようなものについては、修正を施したうえで採用したものもある。以下に一例を示す。

原 文	A computer implemented method, apparatus, and computer program product for generating neuropsychiatric diagnoses.
翻訳文	神経精神病の診断を生成するためのコンピュータ実装方法、装置及びコンピュータ・プログラム製品 を提供する 。

上例の原文は名詞句であるが、翻訳文は語を補って文形式で訳しており、対応が不正確である。ただし、名詞句部分は正確に訳されており、不要な「を提供する」を翻訳文から削除することで正確な対応となるため、これを削除したうえでテストセットに採用した。

なお、原文と翻訳文は、英語と日本語の言語特性の違いにより文の構造等を適切な形に言い換えている場合もある。こうした言い換えは不可避のものであり、原文の内容を正確に伝達していれば採用とした。以下に一例を示す。

原 文	Continuous operation of electrolysis cell 10 requires water transport from cathode 16 to anode 12 where it is consumed.
翻訳文	電解セル 10 を継続的に運転するためには、水を消費するアノード 12 にカソード 16 から水を移動させる必要がある。

上例は、原文の文構造に忠実に訳すと「～の継続的な運転は、～水の移動を必要とする。」となる。しかし、日本語文ではこのような無生物主語構文は多用されず、上記翻訳文のように表現されるのが一般的である。このような言い換えを不対応と見なしてしまうと、特定の文体の原文がことごとく不採用となるおそれがあるため、翻訳文から原文の意味が正確に把握できる限りは不問とした。

[機械翻訳結果の利用]

原文と翻訳文との齟齬は、目視で両者を見比べるだけでは気づかない場合も多い。このため、目視チェックにあたっては、選定した文に対し数種類の機械翻訳結果を補助情報として利用した。機械翻訳結果の利用は、例えば原文におけるスペルミス（⇒未知語となる）や文法誤り（⇒翻訳精度が著しく低下する）、翻訳文における語の補足や省略（⇒対応語が存在しない）等の検出に有用であった。ただし、文の選定時に機械翻訳結果の良い／悪い文を意図的に選定すると、特定の機械翻訳エンジンに対して不当に有利／不利なテストセットになってしまう危険性があるため、その点への留意は必要である。

5.2.3.3. 特許文献の用途上有用である【必須】

特許文献に含まれる文の大部分は技術内容に関するものであり、等しく有用であるが、一部、技術内容に関係しない文も存在する。こうした文は、「特許庁手法」の想定する機械翻訳結果の主要な用途（検索、粗読、精読）においては重要度が低いため、テストセットからは原則除外した。この条件により除外された文の例を下に示す。

原 文	These and various other features and advantages will be apparent from a reading of the following detailed description.
翻訳文	これらの及びさまざまな他の特徴および利点は、以下の詳細な説明を読めば明らかとなるであろう。

5.2.3.4. チェックリスト任意項目の検証に適している【優先】

英日テストセットには、「特許庁手法」の「チェックリストに基づく評価」に定められた各チェック項目の検証に適した文を充分に含めておく必要がある。チェック項目中、【必須項目】に関しては、品詞の誤り、係り受けの誤り等、あらゆる文にて発生する可能性があり、特に意識して文を選定する必要はないが、例えば「数式、化学式に関する翻訳ミス」等の【任意項目】や、「単数形・複数形併記表現を適切に処理していない」等の英日【翻訳言語別任意項目】の検証に適した文は、それぞれ意図的に選定する必要がある。

英日テストセットの文の選定においては、あらかじめこれらのチェック項目の検証に適した文の要件を定めて、これに該当する文を優先的に採用した。検証に適した文の要件とは、例えば「数式、化学式に関する翻訳ミス」というチェック項目であれば「数式又は化学式を含む文」となり、「単数形・複数形併記表現を適切に処理していない」というチェック項目であれば「単数形・複数形併記表現を含む文」となる。以下に一例を示す。

原 文	The electrical machine (102) includes an inner nested stage (444) nested with respect to an outer nested stage (446).
翻訳文	発電機（１０２）が、外側の入れ子段（４４６）に関して入れ子状にされた内側入れ子段（４４４）を含む。

上例の原文は、各要素が付番を含むため、【任意項目】の Z-06「特許特有の記号や番号表記に関する翻訳ミス」の検証要件を、そしてその付番がカッコ書きされているため同 Z-08「原文のカッコ書きに起因する翻訳ミス」の検証要件を満たす。さらに、with respect to という慣用表現を含むため、【翻訳言語別任意項目】の EJ-09「慣用表現とみなすべき場合／そうでない場合を適切に翻訳できていない」の検証要件も満たしている。

5.2.4. チェックリスト項目検証用フラグのセット

5.2.3.4.で説明したとおり、英日テストセットには、【任意項目】及び【翻訳言語別任意項目】の各項目の検証に適した文を優先的に選定しているが、選定した各文に対し、その文により検証可能なチェック項目をあらかじめ記録しておくことにより、テストセットから評価対象文 200 文を選定する際、各チェック項目の検証に適した文を確実に含めることが可能となる。また、評価作業時に、当該文においてチェックすべき項目を評価者に明示することもできるようになる。

英日テストセットでは、この考えに基づき、選定文が【任意項目】及び【翻訳言語別任意項目 [英日]】の検証要件を満たす場合、データ上にフラグをセットして記録した。

表 5-9 及び 5-10 に、英日テストセットにおける【任意項目】及び英日の【翻訳言語別任意項目】の各チェック項目に対応したフラグの設定内容の一覧を示す。

表 5-9. 【任意項目】検証用フラグ設定

チェック項目	フラグ設定内容
Z00 訳ゆれ	1：主要な技術用語が対象 9：その他全ての文
Z01 文の分断	9：全ての文
Z02 主語述語	1：名詞句以外の全ての文
Z03 並列表現	1：典型例 2：並列表現を含む文
Z04 名詞句	1：原文全体が名詞句（タイトル、クレーム等） 2：文中に 4 語以上からなる名詞句を含む
Z05 特許用語表現	1：典型例 2：準典型例（フレーズ等） 3：一般的用語（invention、embodiment 等） 4：claim 9：それ以外の全ての文
Z06 特許記号番号	1：図番号、文献番号等（下記 2 に該当しない記号番号全て） 2：要素付番、クレーム番号
Z07 箇条書き	1：典型例 2：クレーム番号 3：段落番号
Z08 カッコ書き	1：カッコ書きを含む文 2：箇条番号、付番等 3：複数形併記 ※device(s)等 4：数式、化学式のカッコ表記 5：外字対応 ※[alpha]等 6：段落番号 ※[0006]等
Z09 数式化学式	1：数式又は化学式を含む
Z10 数字と単位	1：数字＋単位表現を含む
Z11 数値範囲	1：上下限値が示された数値範囲 ※from 100 to 200 等 2：上限値又は下限値のみ示された数値範囲 ※less than 100 等 3：付番等の数値範囲 ※devices 1－3 等 4：クレーム、図番の数値範囲 ※claims 1 to 10 等

表 5-10. 【翻訳言語別任意項目【英日】】検証用フラグ設定

チェック項目	フラグ設定内容
EJ01 ピリオド	1：文中にピリオドを含む文 2：図番のピリオド ※FIG.1 等
EJ02 コロン	1：コロン(:)の典型例（後続が目的語） 2：コロンを含むが後続が目的語でない 3：コロンの特殊用法 ※SEQ ID:123 等 4：セミコロン(;)の典型例 5：セミコロン(;)だが文末
EJ03 重要複数	1：典型例
EJ04 単複併記	1：XXX and XXXs 型 2：XXX(s)型
EJ05 序数詞	1：不定冠詞を伴う ※a first 等 2：無冠詞 ※first 等 9：定冠詞を伴う ※the first 等
EJ06 語の挿入	挿入された語をセット。
EJ07 倒置文	1：典型例 2：There is provided 型
EJ08 名詞と代名詞	1：典型例 2：代名詞を含む文
EJ09 慣用表現	慣用表現をセット。ただし以下のフラグの場合あり 2：慣用文構造 ※not only, but also 等 9：その他全ての文
EJ10 前置詞	1：典型例 2：前置詞を含む全ての文
EJ11 接頭接尾	該当する接頭辞、接尾辞をセット
EJ12 スラッシュ	1：スラッシュを含むもの（ただし and/or 以外） 2：and/or
EJ13 therein	1：therein、therefor 等 thereXXX 型の語を含む文 2：herein を含む文 3：wherein を含む文 4：whereas 等 whereXXX 型の語を含む文
EJ14 same	1：タイトルで same の代名詞用法を含むもの
EJ15 小文字混入	全文字大文字型のタイトル中に小文字表記を含むもの
EJ16 クレーム引用	1：引用クレーム特殊表現 ※any of the preceding claims 等 2：引用クレーム一般表現 ※of claim 1、according to claim 1 等

【任意項目】【翻訳言語別任意項目】とも、同一チェック項目中で複数の条件に該当する場合は、数字の若いフラグが優先される。例えば【任意項目】の Z-06「特許記号番号」であれば、一文中に特許文献番号（フラグ 1）と要素付番（フラグ 2）が共に含まれていた場合、フラグは 1 がセットされる。

なお、本フラグは、各チェック項目の検証に適した評価対象文をテストセットから選出する用途に加え、評価者に対し、当該文で評価すべきチェック項目を明示する用途にも使用される。具体的には、評価作業のための「品質評価作業用ファイル」において、フラグのセットされていないチェック項目（すなわち評価対象外である項目）は、評価値入力欄をグレイアウトする（6.3.3.8.参照）。

この用途を想定し、各チェック項目において「特に検証に適した文ではないが、該当の可能性がある」文、端的に言えば「品質評価作業用ファイル」上でグレイアウトすべきでない文には、フラグ 9 を立てている。例えば【任意項目】の Z-00「訳ゆれ」では、文中に主要な技術用語が複数回出現する場合にフラグ 1 を立てているが、訳ゆれは主要な技術用語以外にも発生する可能性があるため、他の文においても「訳ゆれ」の評価欄をグレイアウトさせないように、残る全ての文にフラグ 9 を立てている。

5.2.4.1. 【任意項目】検証用フラグ設定

以下【任意項目】各チェック項目の検証用フラグ設定の詳細について説明する。

[Z-00 訳ゆれ]

文中、同一の技術用語が複数回出現する場合にフラグ 1 を立てた。フラグ 1 の文を評価対象文とすることで、訳ゆれの有無を検証することができる。

原 文	フラグ
An electrical test instrument as set forth in claim 1, wherein said second communication mode is an optical communication mode .	1

上例の原文は、communication mode という技術用語が 2 回出現しているためフラグ 1 となる。なお、上例では 2 回目の communication mode は optical communication mode で一語であり、この単位では重複しないが、重複が用語の一部であっても、重複部分を同じ訳語で訳すべき場合はフラグを立てた。

なお、訳ゆれは主要な技術用語以外にも発生する可能性があるため、前述のとおり、上記に該当しない文にもフラグ 9 を立て、全ての文において「訳ゆれ」の評価欄が有効となる（すなわちグレイアウトしない）ようにした。

[Z-01 文の分断]

文の分断はあらゆる文において発生する可能性があるため、フラグは全ての文一律で 9 とした。

文の分断は特に長文や文中にピリオド、コロン等を含む場合に発生しやすいが、これら を評価文としたい場合は、各文の文長情報や他のチェック項目のフラグに基づき抽出すれ ばよい。

[Z-02 主語述語]

原文における主語・述語・目的語（他動詞の場合）の相互関係が翻訳結果にて維持でき るか否かをチェックする項目である。全ての文にて発生可能性があるため、「発明の名称」 等の名詞句を除く全ての文にフラグ 1 を設定した。

[Z-03 並列表現]

並列表現は多用され、かつ様々なバリエーションがあるため、このなかで特に英日機械 翻訳において課題となりそうなものを「典型例」としてフラグ 1 を設定し、それ以外の並 列表現（フラグ 2）と区別した。

「典型例」には、例えば以下のような文を選定した。この文では、まず length と size or height とが並列であり共に that form the gaps 342 を受け may be varied に係っている。 さらにこの構造中で size と height が並列で of the projection 340 を受けている複雑な構造 となっている。

原 文	フラグ
The length of the arms 344 and the size or height of the projections 340 that form the gaps 342 may be varied.	1

なお、並列表現に関する翻訳ミスは、並列表現が原文中に存在しない場合は発生する可 能性は皆無であるため、残りの文はフラグを空欄として評価値入力欄をグレイアウトさせ る。

[Z-04 名詞句]

名詞句を文と誤解釈して翻訳するミスをチェックするための項目であるため、「発明の名 称」や「特許請求の範囲」など、文全体が名詞句であるもの全てにフラグ 1 をセットした。 また、多数の語にて構成する複合名詞においても、これを文節と誤解釈する同様の翻訳ミ スが発生しやすいため、4 語以上からなる名詞句を含む文にはフラグ 2 をセットした。

[Z-05 特許用語表現]

特許特有の用語や表現に関する翻訳ミスをチェックする項目であるが、「特許特有の用語・表現」の幅は非常に広く、特許文献から抽出した文である以上、大半の文において該当の可能性がある。ただし、その中でも特に翻訳ミスの原因になりそうな典型例にフラグ 1 を設定した。例えば下例における **a person of ordinary skill in the art**（当業者）等が該当する。

原 文	フラグ
However, a person of ordinary skill in the art will know that these measurements may be increased or decreased according to the technical requirements of the application.	1

また、特許特有のフレーズだが文法的に翻訳の難易度が低いものは、準典型例としてフラグ 2 を設定した。例えば下例の **in further embodiments**（さらなる実施形態において）等が該当する。

原 文	フラグ
In further embodiments , the resin may be a mixture of amorphous and crystalline resins and the temperature may be above the glass transition temperature of the mixture.	2

その他、embodiment、application 等、単語レベルの特許用語はフラグ 3、“claim”のみが特許用語に該当する場合はフラグ 4 とした。また、それ以外の文についてもフラグ 9 をセットし、評価対象外となることを避けた。

[Z-06 特許記号番号]

特許文献番号（U.S.PAT.No.123456 等）や図番（FIG.1 等）など、特許文献にて使用される記号や番号に起因する翻訳ミスをチェックする項目であり、こうした記号・番号を含む文にフラグ 1 を立てている。また、要素への付番（Device 100 等）とクレーム番号（Claim 10 等）についてもフラグ 2 を立て、特定可能とした。

[Z-07 箇条書き]

典型的な箇条書き表現にフラグ 1 を立てている。以下に一例を示す。

原 文	フラグ
The method involves the following steps: (1) microwave treatment, (2) removal of the paraffin by a centrifugation step, (3) Proteinase K digestion, and (4) a heat step to destroy Protease K activity.	1

また、箇条書き表現に準ずるものとして、クレーム文で冒頭にクレーム番号を伴うものにはフラグ 2 を、明細書の文で冒頭に段落番号（[0001] 等）を伴うものにはフラグ 3 を、それぞれ設定した。

[Z-08 カッコ書き]

カッコ書きに起因する翻訳ミスをチェックする項目であり、カッコ書きを含む全ての文にフラグを立てた。ただし、箇条書き番号や付番に用いられるカッコ表記、device(s)など複数形併記に用いられるカッコ表記、数式や化学式に含まれるカッコ表記には、カッコ表記の一種ではあるがチェック項目の想定する翻訳ミスとは関連が薄いため、それぞれフラグ 2、3、4 をセットして区別可能とした。また、[alpha] 等、半角英字で表現できない文字が置換された角カッコ表記と、[0006] 等の段落番号に用いられる角カッコ表記についても、フラグ 5、6 をそれぞれ設定した。

[Z-09 数式化学式]

文中または文の末尾に数式や化学式を含む文にフラグ 1 を設定した。

原 文	フラグ
At step 507 the node calculates an exponentially-weighted average called a smoothed round-trip time (SRTT) according to the following formula: $SRTT_{new}=[\alpha]RTT_{new}+(1-[\alpha])SRTT_{old}$	1

[Z-10 数字と単位]

文中に数字＋単位（%、ml、centimeter 等）を含む文にフラグ 1 を設定した。

原 文	フラグ
[0035] The blending amount of the above additive is usually not more than 10 parts by weight, preferably 5 parts by weight based on 100 parts by weight of the cross-linkable polymer.	1

[Z-11 数値範囲]

数値範囲を含む文にフラグを設定した。上限と下限の双方の数値が示されている場合はフラグ 1 とした。

原 文	フラグ
[0023] In the highly functional polyethylene fiber of the present invention, a tensile modulus preferably ranges from 200 cN/dtex to 750 cN/dtex.	1

一方、more than や less than、equal to 等、上限・下限等一方の数値のみが示されている場合はフラグ 2 とした。

原 文	フラグ
The method, as recited in claim 5, wherein the photoresist layer is formed from 248 nm photoresist and the feature has a CD not greater than 140 nm.	2

ただし、one or more、two or more 等は数値範囲ではなく慣用表現の範疇であるため、フラグは設定していない。

なお本項目では、上記以外にも付番の範囲（例：devices 1 to 3）にはフラグ 3、クレーム番号、図番の範囲（claims 1 to 10、FIGs.1-4 等）にはフラグ 4 を設定している。

5.2.4.2. 【翻訳言語別任意項目【英日】】検証用フラグ設定

一方、英日【翻訳言語別任意項目】の各チェック項目の検証用フラグ設定は以下の通りである。

[EJ-01 ピリオド]

文中にピリオドを含む文にフラグ 1 を設定している。当然ながら文末のピリオドは対象外である。

原 文	フラグ
The high failure rate caused the inventor to conclude that U.S. Pat. No. 5,628,127 was unpractical for certain applications and caused the inventor to search for improvements.	1

小数点（0.001 等）、一般的な略号（e.g.、i.e.等）も念のためフラグ対象とした。また、図番（FIG.1 等）は数が多いためフラグ 2 として、他と区別可能とした。

[EJ-02 コロン]

文がコロン (:) またはセミコロン (;) を含む文にフラグを設定した。このチェック項目は、機械翻訳結果がコロン前後の係り受け関係を適切に表現できているかをチェックするためのものであるため、下例のように、文がコロンを含み、かつコロンの前後が文として連続しているものを典型例としてフラグ 1 とした。

原 文	フラグ
2. The solid cleaning product of claim 1, further comprising: (b) at least one anionic surfactant; (c) at least one nonionic surfactant; and (d) at least one builder.	1

一方、コロンを含むが、その前後が文として連続しておらず、特に前後の係り受けを意識する必要がない文については、フラグ 2 を設定した。

原 文	フラグ
The method involves the following steps: (1) microwave treatment, (2) removal of the paraffin by a centrifugation step, (3) Proteinase K digestion, and (4) a heat step to destroy Protease K activity.	2

上例は、コロン直前の **steps** までで文として完結しており、翻訳時にコロン以降との係り受けに配慮する必要がない。このため本チェック項目に抵触する可能性は低い。例えばコロンを含む文は必ず後続文との係り受けを配慮するロジックが機械翻訳エンジンに組み込まれているような場合、逆にこのようなケースで誤動作を生じる可能性は皆無ではない。なお同様の理由で、コロンが文区切り以外で用いられている場合（例えば 3:1 等）にもフラグ 3 を設定した。

一方、セミコロン (;) の典型例にはフラグ 4 を設定した。ただしセミコロンは、通常はコロンと共に起するため、より若いフラグ 1 が優先されていることが多い。

原 文	フラグ
In some other embodiments, the emulsions of the present invention comprise a first emulsion emulsified within a second emulsion, wherein (a) the first emulsion comprises (i) an aqueous phase; and (ii) an oil phase comprising an oil and an organic solvent; and (iii) a surfactant; and (b) the second emulsion comprises (i) an aqueous phase; and (ii) an oil phase comprising an oil and a cationic containing compound; and (iii) a surfactant.	4

また、特許文献では、セミコロン部分で強制改行が挿入される場合がある。この場合、その単位が機械翻訳結果の翻訳単位となることから、テストセットにもこのパターンを含めている。こうしたデータは末尾がセミコロンとなる。この場合、フラグ 5 を設定した。

原 文	フラグ
A) providing an ink jet printer that is responsive to digital data signals;	5

[EJ-03 重要複数]

本チェック項目は、原文において名詞の単数形／複数形に大きな技術的意味を持たせている場合に、それが翻訳結果から把握できるかどうかをチェックする項目である。文がこれに該当する典型例であると判断される場合、フラグ 1 をセットした。

なお、テストセットは単文単位であり、文中の単数形／複数形が大きな技術的意味を有するか否かを文脈から判断することはできないため、チェック対象とするかどうかは人手翻訳の内容（単数又は複数であることが強調されているか否か）に依存している。

原 文	人手翻訳	フラグ
The top chamber 38 includes outlet ports 40 and an orifice 42 defined by the valve seat 44.	上部チャンバ 38 は、 複数の出口ポート 40 と、バルブシート 44 により定められる 一つ の オリフィス 42 とを含む。	1

[EJ-04 単複併記]

英語では、要素が単数複数どちらの可能性もある場合や、いずれでも可である場合に、例えば device and/or devices、または device(s) という形でこれを表現することがあるが、日本語には単数複数の概念がないため、英日機械翻訳でこの部分が適切に訳出されない可能性がある。本項目はこれをチェックするためのものであり、検証用の文として、前者のパターンを含む文にフラグ 1、後者のパターンを含む文にはフラグ 2 を設定した。

原 文	フラグ
As noted above, interconnect test module 150 may send a request or other signal to operating system 132 to cause the component or components to be de-allocated.	1

[EJ-05 序数詞]

特許文献では、通常は定冠詞を伴う序数詞が、不定冠詞を伴ったり（例えば a first electrode）無冠詞であったり（first electrode）する人が多い。本項目では、前者に該当する場合はフラグ 1、後者に該当する場合はフラグ 2 をセットした。

原 文	フラグ
For example, a hub-formatted message from a tool may be reformatted or translated to comply with communications data/rules for a first game platform and also for a second game platform.	1

なお、定冠詞を伴う通常の序数詞表現（the first electrode）は、それが原因で翻訳ミスが生じる可能性は低いが、念のため評価対象には加えておくべきであるため、フラグ 9 をセットしてある。

[EJ-06 語の挿入]

複数の語からなる慣用表現の中間への語の挿入が原因の翻訳ミスを検証するための項目であり、典型例についてフラグをセットしてある。なおフラグは、挿入された語そのものを使用している。

下例においては、慣用的表現 deposit XXX upon YYY（YYY に XXX を堆積する）の中間に、本来は後置されるべき as a layer 220 という語句が挿入されている。このようなケースは本項目の典型例と見なし、フラグ欄に挿入語句 as a layer がそのままセットされる。

原 文	フラグ
For example, electroplating or electroless plating could be used to deposit the desired thickness of barrier material as a layer 220 upon container 201.	as a layer

ただし、for example や e.g. など、慣用的に挿入される語句が挿入された文については、典型例とはみなさず、フラグは設定していない。

[EJ-07 倒置文]

文が倒置文である場合にフラグ 1 をセットしている。

原 文	フラグ
Below node 440 is node 420 with a size of 2 and connected by link 520, and node 421 with a size of 2 and connected by link 521.	1

ただし、英語特許文献において多用され、慣用表現ともみなせる **There is provided** 関連の文は、フラグ 2 として他と区別した。

原 文	フラグ
For example, in accordance with the present invention, there is provided a method of increasing embryo size and/or weight.	2

[EJ-08 名詞と代名詞]

原文が名詞とこれを参照する代名詞を共に含む場合、機械翻訳結果において両者の出現順が逆となり、両者間の対応が不明瞭になることがある。この検証に適した文に、フラグ 1 を設定した。

例えば下例では、代名詞 **their** は先行する **gas turbine engines** を指すが、機械翻訳結果においては、最初に **in order** 節が配置され「**これらの**効率を改善するために、より高い**ガスタービンエンジン**の動作温度が～」等と訳される可能性がある。

原 文	フラグ
Higher operating temperatures for gas turbine engines are continuously being sought in order to improve their efficiency.	1

なお、普通に翻訳する限りこうした出現順の逆転は生じないと考えられる文であっても、訳し方によっては逆転が生じる可能性は皆無ではない。したがって、そのような上記典型例に該当しない文であっても、名詞と代名詞を含むものには全てフラグ 2 を設定し、評価対象に加えておくこととした。

原 文	フラグ
Once the system is filled with solvent , it can be mixed by an injection of a gas 6.	2

[EJ-09 慣用表現]

慣用表現が適切に翻訳されているか、及び、慣用表現と見なすべきでない表現が慣用表現として扱われていないかをチェックするための項目であり、文中に含まれる主要な慣用表現をフラグとしてセットしている。

原 文	フラグ
The insulation in contact with the wall of the conduit (see row 5 in Table 1) was assumed to have the properties of ZEDCOR 98 alumina.	in contact with

また、文全体が慣用的な構造になっているものについては、フラグ 2 を設定した。例えば下例は、「the 比較級～、the 比較級～」で「～であるほど、～である」という慣用的な文構造であり、これに該当する。

原 文	フラグ
Generally speaking, the higher is the difference between the dielectric constants, the more significant is the internal reflection.	2

ただし、英文においてどこまでを慣用表現と見なすかは定義が困難であり、また種々の機械翻訳エンジンにおいて、どのような表現が慣用表現と見なされるかは予測不可能である。このため、上記に該当しない文にも全てフラグ 9 を設定し、評価対象に加えておくこととした。

[EJ-10 前置詞]

前置詞は、関連する動詞や目的語により適切に訳し分ける必要がある。本項目は、こうした前置詞の訳語が適正であるかを評価するためのものである。

前置詞の訳語は、例えば of であれば「～の」、for であれば「～のための」等の定番訳が存在するが、本項目の検証のための典型例は、このような定番訳では誤訳となるようなものを選定し、フラグ 1 を設定している。例えば下例における on は、定番訳は「～の上で」等であるが、ここでは feed on という慣用表現の一部となっており、「～を（与えられた）」等と訳出する必要がある。

原 文	フラグ
In particular, the royal jelly secreted by the young worker bees fed on the HDACi may have a changed ratio of 68 to 64 kDa protein of MRJP3.	1

ただし、前置詞の誤訳は、こうした典型的な例でなくとも発生し得るものであるため、前置詞を含む全ての文にフラグ 2 を設定した。

[EJ-11 接頭接尾]

接頭辞、接尾辞を伴う語が含まれている全ての文について、その接頭辞、接尾辞をフラグとしてセットしている。

原 文	フラグ
Non-Sweet Binder Compositions and Methods of Making and Using Same	non

[EJ-12 スラッシュ]

スラッシュを含む全ての文にフラグ 1 を設定している。ただし、頻出語である and/or は、通常のスラッシュ表現と区別し、フラグ 2 とした。

原 文	フラグ
Preferably, the effect(s) of the light chain component is/are reversible.	1

[EJ-13 therein]

英語特許文献で多用される therein、therefor、herein、wherein 等を含む文を対象とし、therein、therefor 等 thereXXX 系の語はフラグ 1、herein は 2、wherein は 3、whereas 等 wherein 以外の whereXXX 系の語は 4 をセットした。

[EJ-14 same]

「発明の名称」にて、代名詞 same を含むものにフラグ 1 をセットした。形容詞用法は対象外である。

原 文	フラグ
WORKFLOW MANAGEMENT APPARATUS, AND METHOD AND PROGRAM FOR THE SAME	1

[EJ-15 小文字混入]

全文字大文字表記の「発明の名称」にて、数値単位や元素記号等を表現するため小文字が使用されているものにフラグ 1 を設定した。

原 文	フラグ
ABSORBABLE STENT HAVING A COATING FOR CONTROLLING DEGRADATION OF THE STENT AND MAINTAINING pH NEUTRALITY	1

[EJ-16 クレーム引用]

「特許請求の範囲」にて、先行する請求項の引用表現を含むものにフラグを設定した。
フラグ 1 は翻訳の難易度が比較的高いものとし、単純な構造の引用表現（of claim 1 や according to claim 1 等）はフラグ 2 とした。

原 文	フラグ
A combined X-ray diffraction and X-ray fluorescence apparatus according to any preceding claim , comprising a plurality of XRD ports.	1

5.2.5. 重要技術用語の特定

最後に、選定した各文に対し、「特許庁手法」の「(2) 重要技術用語の翻訳精度の評価」の対象とすべき重要技術用語を最大 3 語⁵選定した。以下、主な選定基準について記す。

5.2.5.1. 選定対象外とする語

技術内容に関係する用語のみを選定した。つまり、「patent application」や「embodiment」等の特許用語や、企業名等、技術的に有意でない固有名詞は対象としていない。

原 文	Various embodiments of toothbrushes formed in accordance with the principles of the invention have been described above.
翻訳文	本発明の原理 に従って形成される歯ブラシの様々な 実施形態 が上述されてきた。
採用語	①toothbrush（歯ブラシ）

上例の embodiment（実施形態）及び principles of the invention（本発明の原理）は、特許用語であり技術用語ではないため「重要技術用語」には選定しない。

5.2.5.2. 名詞以外の語の選定可否

選定する語は名詞（複合語を含む）を中心としたが、必要に応じて専門的な動詞、形容詞、副詞（例えば decolorize、symmetric 等）も含めた。

原 文	The sliding support units 300 may be provided symmetrically around a circumference of a pressure vessel as shown in FIGS. 2 and 3.
翻訳文	図 2 及び図 3 に示されるように、滑り支持装置 300 は圧力容器の周囲に沿って 対称に 設けられてもよい。
採用語	①sliding support unit（滑り支持装置）、② symmetrically（対称に） 、③pressure vessel（圧力容器）

上例の symmetrically は副詞であるが、技術的に重要な意味を表す語であると判断し、重要技術用語のひとつとして選定した。

⁵ 4.3.2.4.参照。

5.2.5.3. 一般的な単語の選定可否

技術用語であっても、極めて一般的な単語、例えば **sensor**、**controller** 等が、一般的な訳語（「センサ」「コントローラ」等）で訳されているようなものは、現在の英日機械翻訳の水準から考えて、ほぼ正しく訳されることが期待されるため、評価作業負担軽減の観点から、他に候補がない場合も、原則、対象外とした。

原 文	The signals 50 from the sensor 44 are provided to the data acquisition and analysis system 42.
翻訳文	センサ 44 からの信号 50 は、データ取得解析システム 42 に供給される。
採用語	①data acquisition and analysis system（データ取得解析システム）

上例の原文には技術要素として **signals**、**sensor** が存在するが、いずれも一般的な単語であり、対応する訳語も「信号」「センサ」と一般的なものであるため、重要技術用語としての採用は見合わせた。

5.2.5.4. 並列表現等により省略された語の選定可否

並列表現等により複合名詞の一部が省略されている場合、語を補って抽出することは避け、可能かつ妥当な場合は形容詞等で採用した。

原 文	The electrodes can further include float yarns and can be configured in a textured or ribbed construction .
翻訳文	該電極は、さらに、浮き撚り糸を有し、 織物構造またはうね織り構造 の中に組み込まれることができる。
採用語	①float yarn（浮き撚り糸）、② textured （織物の）、③ ribbed construction （うね織り構造）

上例原文の **textured or ribbed construction** は、翻訳文のとおり **textured construction**（織物構造）と **ribbed construction**（うね織り構造）の意味であり、前者の **construction** が省略された形となっている。このような場合、省略語を補って **textured construction** を抽出するようなことはせず、原文どおり形容詞 **textured** として採用した。

5.2.5.5. 長大な複合語の選定可否

複合語はそのまま採用することを原則としたが、過度に長大である等、1用語として扱うのが適当でないと判断される場合は、形容詞での採用等、適切な単位で抽出した。

原 文	In certain embodiments, Br ₂ can be a distilled, anhydrous, acid-free reagent .
翻訳文	ある実施形態において、Br ₂ は 蒸留された、無水の酸不含試薬 であることができる。
採用語	① distilled (蒸留された)、② anhydrous (無水の)、③ acid-free reagent (酸不含試薬)

原文の **distilled, anhydrous, acid-free reagent** (蒸留された、無水の酸不含試薬) は 1 つの複合語ともみなせるが、1 用語として扱うには長大すぎると判断し、**distilled** と **anhydrous** はそれぞれ形容詞として採用し、**acid-free reagent** のみ名詞で採用した。

5.2.5.6. 選定語の整形

選定した語は、名詞は単数形、動詞は原形に修正して抽出した。動詞は訳語も終止形とした。なお、翻訳文において意識されている（例えば他の品詞に言い換えられている）場合は、原語の品詞に応じて対訳語を修正した。

原 文	Continuous operation of electrolysis cell 10 requires water transport from cathode 16 to anode 12 where it is consumed.
翻訳文	電解セル 10 を継続的に運転するためには、水を消費するアノード 12 にカソード 16 から 水を移動させる 必要がある。
採用語	①electrolysis cell (電解セル)、② water transport (水を移動させること)、③cathode (カソード)

上例原文の **water transport** は名詞だが、翻訳文では「水を移動させる」と動詞で表現されている。選定にあたっては、これを原文の品詞に合わせて「水を移動させること」と名詞形に修正した。

5.3. 英日テストセットのフォーマット

本項に英日テストセットのデータフォーマットを示す。

なお、英日テストセットは、そのまま評価作業に使用するのではなく、ここから評価対象文を選定して別ファイルに保存し、評価対象とする各機械翻訳システムで作成した機械翻訳結果をこのファイルに追加したうえで、これを評価作業用ファイルに編集するという手順を踏む。この手順の詳細は次章「英日機械翻訳の実施」を参照のこと。

5.3.1. 英日テストセットのファイル形式とデータ構成

英日テストセットは、Excel ファイル形式で作成する。データは、以下のとおり構成される。

表 5-11. 英日テストセット データ構成

データ種別	Excel 行	解 説
見出し行	1 行目	データ行の各項目の内容を示す行。
パラメータ行	2 行目	評価作業用ファイル作成時に各項目の有効無効等をコントロールするための行。
データ行	3 行目～ 1,002 行目	テストセットに選定した文とその翻訳文、重要技術用語、検証フラグ等から成る。

次節にてデータ行の構成を説明する。なお、パラメータ行については次章（6.3.3.2.）で説明する。

5.3.2. データ行の構成

英日テストセットの各データ行は、以下の構成となっている。

表 5-12. 英日テストセット データ行構成

データ項目	Excel 列	入力値	解 説
空欄	A	(空欄)	評価作業対象文選定時などに任意に使用できる欄。
文番号	B	A001～H212	IPC セクション別の連番。
原文献番号	C	原文献の文献番号	米国文献は USyyyynnnnnn、欧州文献は EPnnnnnnnn。
対応文献番号	D	対応文献の文献番号	JPyyyynnnnnn。
記載箇所	E	Title / Abst / Claim / Desc	文の抽出箇所（「発明の名称」「要約」「特許請求の範囲」「発明の詳細な説明」）を示す。
IPC	F	IPC サブクラス	筆頭 IPC のサブクラス。
原文	G	原文データ	英語原文。
語数／文字数	H	原文の語数	英日原文の語数。
人手翻訳	I	人手翻訳データ	対応特許から採取した翻訳文。
機械翻訳 1～10	J～S	(空欄)	評価対象文の選定後に機械翻訳結果をセットするための欄。
サンプル評価	T	○	評価対象文のうち、「流暢さの評価」の対象とする文を特定するための欄。本欄が空欄の場合、評価欄がグレイアウトされる。
重要技術用語①～③	U,W,Y	重要技術用語	文から選定した重要技術用語。選定した用語数により空欄の場合あり。
(用語①～③の対訳)	V,X,Z	重要技術用語に対応する人手翻訳	重要技術用語①～③に対応する人手翻訳中の訳語。
チェック項目検証フラグ	AA～	検証用フラグ	【任意項目】及び【翻訳言語別任意項目】の各項目に対応した検証フラグ。必要に応じて列の追加・削除や順序の変更が可能。

英日テストセットのデータサンプルは、添付資料 5-1 を参照のこと。

「原文」欄に関する注記

特許文は、公報における読みやすさ等のため、文中のコロン、セミコロン等にて意図的に改行されている場合があるが、英日テストセットでは、文全体を対象とするため、これらの改行を除去したうえで全文を 1 セル中に納めている。

第 6 章

英日機械翻訳の実施

第 6 章 英日機械翻訳の実施

英日テストセットから評価対象文 200 文を抽出し、5 種類の英日機械翻訳エンジンを用いて機械翻訳結果を取得して、「特許庁手法」による品質評価を実践するための「品質評価対象データ」を作成した。以下、その内容について記す。

6.1. 評価対象文の選定

品質評価対象データを作成するための評価対象文は、以下に示す「特許庁手法」の定義に従い 200 文とした。

2. テストセットから評価対象件数の機械翻訳対象文を抽出する（評価対象件数は 200 文程度）。機械翻訳対象文の抽出は、テストセットの作成と同様、技術分野等について偏りないように行うことを原則とする。ただし、評価の目的や評価作業のリソースに配慮した調整は適宜行ってよい。

評価対象文 200 文における「記載箇所」「技術分野（IPC）」別の構成比率も、上記定義に従い、英日テストセットにおけるそれぞれの構成比率に準拠させた。ただし「原文の長さ」に関しては、評価作業負担の軽減のため、長文の比率を最小限とするよう調整した。

また、評価対象文 200 文には、「チェックリストに基づく評価」の【任意項目】(4.3.6.) 及び【翻訳言語別任意項目 [英日]】(4.3.7.) の各チェック項目の検証に適した文を充分に含めた。具体的には、英日テストセットの「検証用フラグ (5.2.4.)」を参照して、各チェック項目の検証に有用な文を複数文選定した。

以下、上記した評価対象文 200 文の選定条件それぞれの詳細を示す。

6.1.1. 記載箇所

評価対象文 200 文における記載箇所別構成比は、実際の特許文献（サンプル 20 文献）における構成比率をベースとして定めた英日テストセットの構成比（第 5 章 5.2.2.1.参照）に準拠し、「タイトル」と「要約」を各 5%、「クレーム」を 20%、「明細書」を 70%とした。下表 6-1 に、この構成比により配分される記載箇所別の文数を示す。

表 6-1. 評価対象文 200 文 記載箇所別構成比

	タイトル	要約	クレーム	明細書	合計
構成比	5%	5%	20%	70%	100%
文 数	10 文	10 文	40 文	140 文	200 文

6.1.2. 技術分野（IPC）

次に、技術分野（IPC セクション）別の構成比率を定めた。英日テストセットでは、IPC セクション別構成比を近年の米国及び欧州公開公報における構成比率をベースに定めており（5.2.2.3.参照）、評価対象文 200 文もこれと同一の比率を採用した。

表 6-2 に、この比率にて各記載箇所の文数を IPC セクション別に割り振った結果を示す。

表 6-2. 評価対象文 200 文 IPC セクション別構成比

IPC セクション	A	B	C	D	E	F	G	H	合計
タイトル	2 文	2 文	1 文	1 文	1 文	1 文	2 文	2 文	12 文
要 約	2 文	1 文	1 文	1 文	1 文	1 文	3 文	2 文	12 文
クレーム	6 文	5 文	5 文	1 文	1 文	3 文	10 文	9 文	40 文
明細書	21 文	21 文	18 文	2 文	2 文	10 文	33 文	29 文	136 文
文 数	31 文	29 文	25 文	5 文	5 文	15 文	48 文	42 文	200 文

6.1.3. 原文の長さ

英日テストセットの文長別構成比率は、原文の語数を基準とし、実際の英語特許文献（サンプル 20 文献）における比率をベースとして、1～25 語（短文）を 50%、26～50 語（中文）を 40%、51 語以上（長文）を 10%としている。

評価対象文 200 文も、本来であればこの構成比に従うのが最良であるが、本調査においては、評価者の作業負担が大きい「チェックリストに基づく評価」を、200 文（×5 エンジン）全文に対し、全チェック項目について行う必要があることから（7.1.1.にて後述）、作業負担軽減のため、51 語以上の長文の比率を最小限に抑える調整を行なった。具体的には、長文の比率を 10%→5%に下げ、代わりに中文の比率を 40%→45%とした。したがって、評価対象文 200 文の文長別構成比率は下表 6-3 のとおりとなる。

表 6-3. 評価対象文 200 文 文長別構成比

	1～25 語（短文）	25～50 語（中文）	51 語～（長文）	合計
文 数	100 文	90 文	10 文	200 文
構成比	50%	45%	5%	100%

なお、上記調整により、評価対象文 200 文の構成は、実際の特許文献よりも長文の比率がやや低くなり、翻訳の難易度が低下すると考えられる。しかしながら「特許庁手法」では、「2. 相対評価手法」にて、評価結果を文長別に集計し、それぞれを実際の特許文献に従い重み付けすることを定めており（4.3.8.3.）、これによって適正な難易度のものに補正できる。

6.1.4. その他の構成比率

「文献種別」、すなわち米国文献と欧州文献の比率は、英日テストセットでは88.3%:11.7%であるが、評価対象文 200 文においては他の条件を優先し、特に配慮しなかった。結果、両者の文数は 191 文（95.5%）：9 文（4.5%）と、やや欧州文献の比率が少なくなった。

また「出願時期」「出願人」についても、英日テストセットにおいて 1 文献あたりの採用文数を原則 1～2 文程度としているため、ここからランダムに選択すれば過剰な偏りは生じないものと判断した。

6.1.5. 検証用フラグの確認

評価対象文 200 文は、「特許庁手法」の「チェックリストに基づく評価」における【任意項目】及び【翻訳言語別任意項目 [英日]】の検証に適した文を十分な量含んでいる必要がある。このため、上記 6.1.1.～3.の各構成比率に従い選定した 200 文に対し、各文に設定された検証用フラグ（5.2.4.）を確認し、各チェック項目の検証に適した文が 200 文中に十分な数含まれているかどうかを確認した。

表 6-4～6-5 に、各チェック項目に対して検証適性の観点から有意なフラグ値¹が設定されている文の数を示す。

表 6-4. 【任意項目】有意検証フラグ設定数

	Z00	Z01	Z02	Z03	Z04	Z05	Z06	Z07	Z08	Z09	Z10	Z11
チェック項目	訳ゆれ	文の分断	主語述語	並列表現	名詞句	特許用語表現	特許記号番号	簡条書き	カッコ書き	数式化学式	数字と単位	数値範囲
フラグ	1	—	1	1	1-2	1-2	1-2	1-3	1	1	1	1-2
文数	62	—	149	47	76	20	103	47	18	3	21	19

※「フラグ」欄には、各項目の有意なフラグ値の範囲を示した。

※Z01 は全ての文に一律 9 が設定されているためカウント対象外とした。

¹ 各チェック項目に設定されたフラグには、評価対象外となるのを避けるために設定されたフラグ 9 等、検証適性の観点からは有意といえないものもある。

表 6-5. 【翻訳言語別任意項目〔英日〕】有意検証フラグ設定数

チェック項目	E01	E02	E03	E04	E05	E06	E07	E08	E09	E10	E11	E12	E13	E14	E15	E16
	ピリオド	コロン	重要複数	単複併記	序数詞	語の挿入	倒置文	名詞代名詞	慣用表現	前置詞	接頭接尾	スラッシュ	Therein	Same	小文字混入	クレーム引用
フラグ	1-2	1-5	1	1-2	1-2	—	1-2	1	—	1	—	1	1-4	1	1	1
文数	18	16	5	6	15	25	10	2	73	48	34	14	42	5	1	9

※「フラグ」欄には、各項目の有意なフラグ値の範囲を示した。

※E06、E09、E11 は語句自体がフラグ値につき範囲なし。

上表に示すとおり、大半のチェック項目に対して 5 文以上の検証文が評価対象文 200 文に含まれていることが確認できた。なお、EJ08「名詞代名詞」は 2 文、EJ15「小文字混入」は 1 文と文数が少ないが、前者は「名詞と代名詞が翻訳結果において逆転する現象」という、現状の機械翻訳エンジンにとって難易度の高い課題に関するチェック項目であり、一方後者は「全文字大文字タイトルへの小文字の混入」という、実際の特許文献における発生率自体が少ないケースに関する項目であることから、いずれも適正な文数であると判断した。

したがって、今回抽出した評価対象文 200 文において文の差し替えは必要ない。なお、仮に検証に適した文数が不十分なチェック項目が検出された場合は、当該項目に有意なフラグが立っている文を英日テストセットから抽出し、記載箇所や技術分野の構成比を維持しながら文の差し替えを行なえばよい。

6.2 英日機械翻訳の実施

翻訳対象文 200 文の各文に対し、5 種類の英日機械翻訳エンジンを用いて 5 種類の機械翻訳結果を作成した。以下、その詳細を示す。

6.2.1. 英日機械翻訳エンジン

英日機械翻訳エンジンには、第 3 章「英日機械翻訳の現状分析」で用いた 5 種類 (RBMT1～3、SMT4～5) を引き続き採用した。各エンジンの設定等についても変更していない（詳しくは 3.1.1. 参照）。

6.2.2. 文分割に対する処理

翻訳対象文 200 文は 1 文単位で翻訳したが、機械翻訳処理の結果、文が複数に分割されて出力される場合がある。この場合、翻訳対象文と 1:1 の対応関係とするため、以下の基準で連結を行なった。

6.2.2.1. 文の分割箇所が自明である場合

文の分断箇所が自明である（文末が句点やコロン等で終わる）場合は、そのまま連結した。

原 文	A number of ultrasound catheter devices have been described, for example in patent application Ser. Nos. 10/229,371, 10/345078, 10/375,903, and 10/410,617, which were previously incorporated by reference.
連結前	多くの超音波カテーテル装置は、特許出願サービスに例えば述べられている。 番号 10/229,371(それは以前に参照によって組込まれた)、10/345078、10/375,903 および 10/410,617。
連結後	多くの超音波カテーテル装置は、特許出願サービスに例えば述べられている。番号 10/229,371(それは以前に参照によって組込まれた)、10/345078、10/375,903 および 10/410,617。

上例では、機械翻訳結果（連結前）では文が 2 文に分断されているが、最初の文の末尾が句点（。）であるため、そのまま後続文と連結しても、文が分断されていることは明らかである。このような場合は、そのまま連結した。なお、コロン（:）やセミコロン（;）での改行も、同様の扱いとした。

6.2.2.2. 文の分割箇所が自明でない場合

そのまま連結すると文の分割箇所が不明となる場合は、接続箇所に「///」を挿入し、文の分断が生じていることを明示した。

原 文	In an exemplary method of making solid cleaning products, the method includes the steps of: (i) adding one or more solid cleaning product components to a mixing vessel; and (ii) mixing the one or more solid cleaning product components via a mechanical stirrer for a desired period of time.
連結前	固体洗浄製品を製造する典型的な方法において、方法は、次のステップを含む (i) は、練り舟に一つ以上の固体洗浄製品成分を追加する以下の通り
	そして、所望の期間の間の機械的攪拌器を介して一つ以上の固体洗浄製品成分を混合している (i i)。
連結後	固体洗浄製品を製造する典型的な方法において、方法は、次のステップを含む/// (i) は、練り舟に一つ以上の固体洗浄製品成分を追加する以下の通り///そして、所望の期間の間の機械的攪拌器を介して一つ以上の固体洗浄製品成分を混合している (i i)。

上例は、機械翻訳結果（連結前）は 3 文に分割されているが、各文の末尾に句点等がないため、そのまま連結すると文の分割箇所が分からなくなり、評価結果に不当な影響を及ぼす危険性がある。このため、連結箇所に「///」を挿入し、文の分割箇所を明示した。

6.3. 品質評価対象データの作成

前項 6.2.にて作成した英日機械翻訳結果を用いて、「特許庁手法」に基づく品質評価を実践するための「品質評価対象データ」を作成した。以下にその手順を示す。

6.3.1. 英日テストセットからの文の抽出

英日テストセットから、評価対象文 200 文を抽出した。具体的には、選定した文に対し、英日テストセット（Excel ファイル）の「空欄（A 列）」にチェックマークを入力してこれを特定し、200 文の選定完了後、それ以外の文を削除したうえで別ファイル（「品質評価対象データファイル」）として保存した。なお、英日テストセットの 1～2 行目の「見出し行」及び「パラメータ行」は、後続処理で使用するため削除せず残しておく必要がある。

英日テストセットのフォーマットについては第 5 章 5.3.「英日テストセットのフォーマット」を参照のこと。

6.3.2. 機械翻訳結果のセット

6.3.1.で作成した「品質評価対象データファイル」は、英日テストセットのフォーマットに準拠した Excel ファイルであるため、J～S 列に「機械翻訳 1～10」欄が空欄で存在している。

このうち「機械翻訳 1」欄～同「5」欄に、6.2.にて作成した 5 種類（RBMT1～SMT5）の機械翻訳結果をセットした。

6.3.3. 品質評価作業用ファイルの作成

「品質評価対象データファイル」には、評価結果を入力する欄が存在しない。また、各項目の検証用フラグ等、評価作業時に提示する必要のない情報も含んでいる。このため、これを「品質評価作業用ファイル」に整形する必要がある。以下に、その手順について記す。

6.3.3.1. 品質評価作業用ファイル作成マクロ

「品質評価作業用ファイル（以下「作業用ファイル」という。）」の整形を、評価の都度人手で行うのは非効率であるため、整形専用のマクロファイルを用意した。

このマクロファイルは、「品質評価対象データファイル」を入力データとし、これを作業用ファイルのフォーマットに整形し、さらに人手評価補助ツールとして辞書引き機能等を組み込んで、マクロ入り Excel ファイルとして出力する機能を有する。詳しい使用方法是「品質評価作業用ファイル作成マクロの使用方法.docx」（電子データ）を参照のこと。

6.3.3.2. パラメータ行の使用

「品質評価対象データファイル」は、2 行目に「パラメータ行」を含む (5.3.1.参照)。このパラメータ行は、「品質評価作業用ファイル作成マクロ」を制御するために用いられる。具体的には、以下の 2 点について設定を行う。

- ・評価対象とする機械翻訳欄
- ・評価対象とする任意項目及び翻訳言語別任意項目

〔評価対象とする機械翻訳欄〕

「機械翻訳 1～10」欄 (J～S 列) のうち、作業用ファイルへの出力対象とするか否かのコントロールは、パラメータ行の同欄に何らかの文字が入っているか否かにより行われる。

本調査では「機械翻訳 1～5」欄を出力対象とする必要があるため、パラメータ行の J 欄～N 欄に任意の文字列 (具体的には MT1～MT5) を入力した。なお、入力した文字列は、対応する機械翻訳結果とともに作業用ファイルに表示される。

〔評価対象とする任意項目及び翻訳言語別任意項目〕

チェックリスト評価において、各【任意項目】及び【翻訳言語別任意項目】は、評価目的等に応じて任意に選択されるものであり、選択された項目のカウント欄のみが作業ファイル上に表示されるべきである。このためパラメータ行にて、作業用ファイル上に表示するチェック項目をコントロール可能とした。

具体的には、パラメータ行の、各チェック項目の検証フラグ欄に該当する欄 (AA 欄～) に「○」がセットされている場合のみ、そのチェック項目のカウント欄を作業用ファイル上に表示する。英日テストセットのパラメータ行は、デフォルトで全てのチェック項目に「○」がセット済みなので、ここから不要なチェック項目の○を削除して空欄とすれば、その項目のカウント欄は作業用ファイル上に表示されなくなる。

ただし本調査は全てのチェック項目について評価を行うため、各チェック項目のパラメータ欄はデフォルトのまま (○) としている。

6.3.3.3. サンプル評価欄

作業用ファイル作成に先立ち、「流暢さの評価」のサンプリング評価対象文についても設定する必要がある。

「特許庁手法」では、「(3) 流暢さの評価」は 200 文中 50 文のみのサンプリング評価とすることが定められている。したがって、作業ファイルにて「流暢さの評価」の対象となる文を評価者に明示する必要がある。これをコントロールするのが、英日テストセットの「サンプル評価」欄 (T 列) である。

作業用ファイル整形マクロでは、入力される品質評価対象データファイルの各行のサンプル評価欄を参照し、ここが空欄の場合、「流暢さの評価」の評価結果欄をグレイアウトする。英日テストセットでは、全ての文の「サンプル評価」欄にはデフォルトで「○」が入っているため、整形マクロの実行前に、「流暢さの評価」の対象とする 50 文以外の文の「サンプル評価」欄を空欄にしておく必要がある。

ただし、本調査では、サンプリング評価文数として設定した 50 文の妥当性を検証する必要があるため、200 文全文について「流暢さの評価」を実施する。このため、各文とも「サンプル評価」欄を空欄にする必要はなく、デフォルト値（○）のままにしておけばよい。

※「流暢さの評価」は、品質評価実践による妥当性検証の結果、200 文全文に対して行うべきことが判明したため、サンプリング評価は廃止となった（9.2.3.参照）。したがって、以後「サンプル評価」欄は常時デフォルトのまま（○）でよい。

6.3.3.4. 作業用ファイル整形マクロの実行

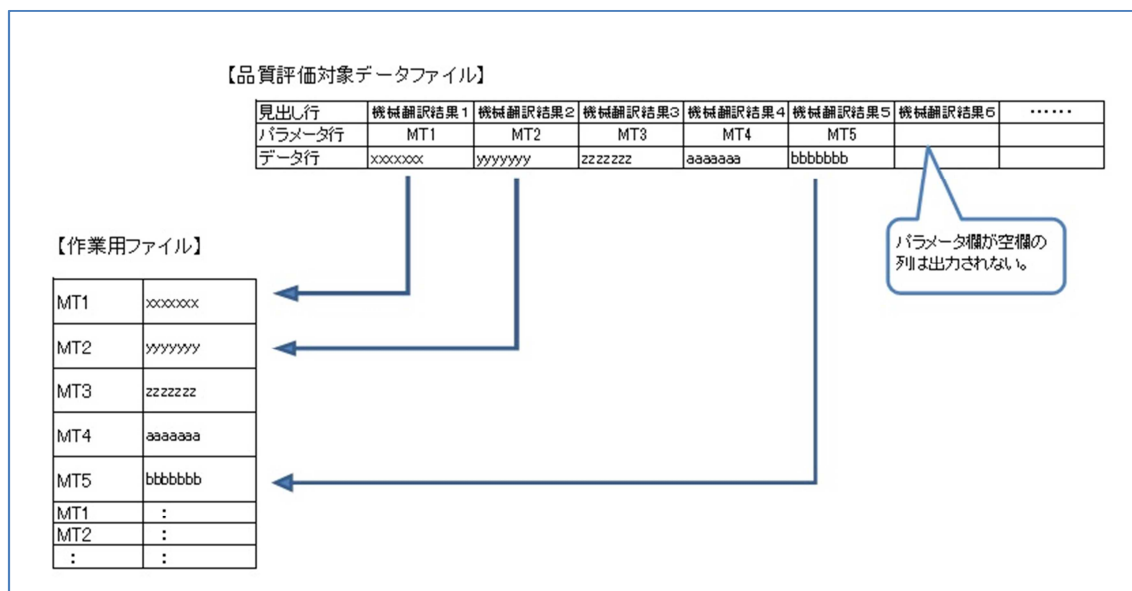
パラメータ行及びサンプル評価欄の設定完了後、作業用ファイル整形マクロを実行する。作業用ファイルは別ファイルとして表示されるので、任意のファイル名にて保存する。なお、作業用ファイルは後述の辞書引き機能を含むため、マクロ入り Excel ファイル（.xlsm）で保存する。また、評価作業は技術分野別を実施する場合、あらかじめ品質評価対象データファイルを IPC セクション別に複数作成しておく必要がある。

【注意】

マクロを実行するため、Excel にて「VBA プロジェクトオブジェクトモデル」を有効にする必要がある。ファイル＞オプション＞セキュリティセンター＞セキュリティセンターの設定＞マクロの設定＞と進み、VBA プロジェクトオブジェクトモデルへのアクセスを信頼する」をチェックする。

6.3.3.5. 作業用ファイルのフォーマット

作業用ファイルは、1 文に対して複数の機械翻訳結果が評価対象となる場合、それぞれの機械翻訳結果を別の行に表示する。本調査のように 5 種類の機械翻訳結果が評価対象となる場合、1 文あたり、5 種類の機械翻訳結果を 5 行に分けて出力する。



また、「特許庁手法」では、「(3) 流暢さの評価」を最初に行うことが定められているため、作業用ファイルの評価値入力欄も「流暢さ」の評価結果入力欄を最初に配置する。

下表 6-6 に、作業用ファイル整形マクロの出力する作業用ファイルのフォーマットを示す。

表 6-6. 作業用ファイルのフォーマット

項 目	Excel 列	内 容
No.	A	作業用ファイル先頭行からの連番。
文番号	B	IPC 別「品質評価対象データファイル」（※以下「入力ファイル」という。レイアウトは「英日テストセット（表 5-12）」と同一）の「文番号（B 列）」欄の値。
文献番号	C	入力ファイルの「原文文献番号（C 列）」欄の値。
対応番号	D	入力ファイルの「対応文献番号（D 列）」欄の値。
記載箇所	E	入力ファイルの「記載箇所（E 列）」欄の値（Title/Abst/Claim/Desc）。
IPC	F	入力ファイルの「IPC（F 列）」欄の値（筆頭 IPC サブクラス）。
原文	G	入力ファイルの「原文データ（G 列）」欄の値。
語数／文字数	H	入力ファイルの「語数／文字数（H 列）」欄の値（原文の語数）。

人手翻訳	I	入力ファイルの「人手翻訳 (I 列)」欄の値。
機械翻訳 ID	J	後続「機械翻訳結果」欄にセットされる機械翻訳結果に対応する機械翻訳 ID として、入力ファイルのパラメータ行に入力された文字列をセット。
機械翻訳結果	K	入力ファイルの「機械翻訳 1～10 (J～S)」欄のうち、パラメータ行で出力対象となったものを、1 行につき 1 つずつセット。つまり複数の機械翻訳欄が出力対象となった場合は、入力ファイル 1 行に対し複数行が出力される。なおその場合、本欄と「機械翻訳 ID 欄」及び「No.」欄以外の各欄は同一の内容がセットされる。
メモ欄	L	空欄。評価作業時の備忘録や評価根拠のコメント等、任意に使用可能。
流暢さ 1-5	M	「(3) 流暢さの評価」の評価値 (5～1) を入力するための空欄。なお入力ファイルの「サンプル評価 (T 列)」欄が空欄の場合はグレイアウトされる。
内容伝達 1-5	N	「(1) 内容の伝達レベルの評価」の評価値 (5～1) を入力するための空欄。
重要技術用語①～③	O,R,U	入力ファイルの「重要技術用語①～③ (U, W, Y 列)」欄の値をセット。選定された用語数により空欄の場合あり。
(用語①～③の対訳)	P,S,V	重要技術用語①～③の訳語。入力ファイルの U, W, Y 列の値をセット。
用語①～③評価 A-D	Q,T,W	「(2) 重要技術用語の翻訳精度の評価」の評価値 (A～D) を入力するための空欄。
重要技術用語数 (※非表示)	X	各行における重要技術用語の語数 (0～3 語) が自動計算されセットされる。評価値集計用の情報であり評価作業時は不要につき、非表示設定となっている。
【必須項目】A～G カウント欄	Y～AE	「(4) チェックリストに基づく評価」の【必須項目】A～G それぞれのカウント値を入力するための空欄。
A～G 合計	AF	上記【必須項目】A～G カウント欄の合計を自動計算し表示する行。
【任意項目】及び【翻訳言語別任意項目】カウント欄	AG～	「(4) チェックリストに基づく評価」の【任意項目】及び【翻訳言語別任意項目】のうち、入力ファイルのパラメータ行で評価対象とされた (=「○」がセットされた) 項目について、カウント値を入力するための空欄が表示される。各欄とも、対応する検証フラグが空欄である文においてはグレイアウトされる (ただしカウント値入力可能)。なお、各欄の見出しは入力ファイルの見出し行の内容が使用される。

作業用ファイルのデータサンプルは、添付資料 6-1「品質評価作業用ファイル データサンプル」を参照のこと。

6.3.3.6. 重要技術用語のハイライト

「原文」欄（G 列）においては「重要技術用語①～③」（P, S, V 列）と一致する語を、「人手翻訳」欄（I 列）及び「機械翻訳結果」欄（K 列）においてはその対訳語①～③（Q, R, S 欄）と一致する語を、欄中赤字でハイライト表示する。

なお、「重要技術用語」は、名詞であれば単数形、動詞であれば現在形でセットされているため、「原文」中にこれとマッチしない活用形（例えば名詞の複数形 `body→bodies` や動詞の活用形 `study→studies, studied` 等）で出現している場合はハイライトされない。また、技術用語の対訳語も、セットの際に例えば動詞を終止形とする等の整形を施しているため、「人手翻訳」欄や「機械翻訳結果」欄にこれと異なる形で出現している場合はハイライトされない。

6.3.3.7. 辞書引き機能

作業ファイルには、「原文」欄をダブルクリックすると、その文に含まれる全ての用語の辞書引き結果を一覧表示する機能が（辞書データファイルがサーバ上に提供される前提で）組み込まれている。

本機能の詳細は、「品質評価作業用ファイル作成マクロの使用法.docx」（電子データ）を参照のこと。

6.3.3.8. チェックリスト項目のグレイアウト

作業ファイルの「チェックリスト【任意項目】及び【翻訳言語別任意項目】カウント欄（AG 列～）」にて、評価対象文が該当する可能性のないチェック項目は、あらかじめグレイアウトされる。

例えば、【任意項目】の Z-11「数値範囲に関する翻訳ミス」であれば、このチェック項目に該当する可能性があるのは、原文中に数値範囲が含まれている場合のみであるため、数値範囲を含まない全ての文は、Z-11 のカウント欄があらかじめグレイアウトされる²。こうすることで、確認すべきチェック項目と確認不要なチェック項目とを評価者に明示することができる。

² グレイアウトされたチェック欄にもカウント値の入力は可能。

第 7 章

特許庁機械翻訳品質評価手法の実践

第 7 章 特許庁機械翻訳品質評価手法の実践

第 6 章にて作成した「品質評価対象データ」に対して、「特許庁手法」を用いて実際に品質評価作業を実践し、その評価結果を反映させた「品質評価結果データ」を取得した。以下、その詳細を記す。

7.1. 「特許庁手法」の実践内容

「特許庁手法」を用いた品質評価の実践は、以下の内容にて行なった。

7.1.1. 評価観点及び対象文数

「特許庁手法」には、下記 4 つの評価観点が定義されている。

- (1) 内容の伝達レベルの評価
- (2) 重要技術用語の翻訳精度の評価
- (3) 流暢さの評価
- (4) チェックリストに基づく評価

本調査では、これら全ての観点による品質評価を実施した。「(4) チェックリストに基づく評価」では、【任意項目】及び【翻訳言語別任意項目】の全項目をチェック対象とした。

評価対象文数は、「特許庁手法」では「(3) 流暢さの評価」は 50 文のサンプル調査、「(4) チェックリストに基づく評価」は任意設定と定められているが、今回の品質評価実践では、全ての観点にて品質評価対象データ全件（200 文）を評価対象とした。

7.1.2. 評価者

「特許庁手法」では、各評価観点における評価者の要件を以下のとおり定義している。

表 7-1. 「特許庁手法」評価者要件定義

評価観点	評価者の要件
(1) 内容伝達レベル	・ 評価対象件数について、複数人（3 人以上）による重複評価を行う。 ・ 評価対象となる各案件の属する各技術分野に精通した評価者が行う。
(2) 重要技術用語	・ 評価対象となる各案件の属する各技術分野に精通した評価者が行う。
(3) 流暢さ	・ 評価対象となる各案件の属する各技術分野に精通した評価者が行う。
(4) チェックリスト	・ 機械翻訳対象文に用いられる言語の読解力を備え、かつ評価対象となる各案件の属する各技術分野に精通した評価者が行う。

この定義を踏まえ、本調査の品質評価実践には、下記の評価者を起用した。

【評価者 A：英日特許翻訳者】

英日間の翻訳を業とする特許翻訳者を分野別に起用し、(1)～(4)の全ての観点の品質評価を実施した。

【評価者 B：機械翻訳専門家】

機械翻訳の研究・開発を業とする技術者を起用し、「(1) 内容の伝達レベルの評価」の重複評価を実施した。

【評価者 C：和文抄録校閲者】

米国和文抄録校閲者を分野別に起用し、「(1) 内容の伝達レベルの評価」の重複評価を実施した。

7.1.3. 評価観点及び評価基準の周知

7.1.3.1. 評価手順書の配布

「特許庁手法」の意図するとおりの観点及び基準にて品質評価が行われるよう、評価の解説と実例を収録した「評価手順書」を各評価観点について作成し、評価者に配布した。

※「評価手順書」の内容は添付資料 7-1～7-6 を参照のこと。

7.1.3.2. 評価者間の水準合わせ

「(1) 内容の伝達レベルの評価」は、各文に対して複数の評価者による重複評価が行われる。このため、「特許庁手法」では、事前にサンプル評価を実施して、評価者間の意識合わせを行うことが定められている。本調査もこれに従い、評価者 A～C 間にて評価水準を合わせるためのサンプル評価を実施した。

サンプル評価は、理想的には、全評価者が同一のサンプル文に対して評価を行い、その結果を照合しながら評価者間で最適の評価水準を定めることが望ましい。

しかしながら、本調査では、評価者の体制やスケジュール等の事情によりこれを実施することが困難であったため、代替策として、各評価者の 1 日分の評価結果（対象文は任意）を回収し、これをあらかじめ準備したリファレンス評価の結果と照合、その結果双方の評価値に 2 以上の相違があるものに対し、リファレンス評価の評価理由等のコメントを添えて再確認を促す形とした。

フィードバックの実例を以下に示す。

原 文	In an embodiment, the inert fluorescent tracer is fed to the feedwater in known proportion to the scaling and/or biofouling agent.				
人手翻訳	実施形態の一つとして、不活性蛍光トレーサは、スケーリング及び／又は 生物付着剤 に対して 既知の割合 で、給水へ供給される。				
機械翻訳	実施形態において、不活性蛍光トレーサは、 既知の スケーリング及び／又は エージェント に 比例 して現れるため、給水に供給される。				
評価者の 評価結果	4	リファレンス 評価結果	2	コメント	「生物付着」という技術用語が欠落している点、「既知」が比例に係っていない点等を重大な誤りと判断しました。

なお、本調査における品質評価実践の主目的は「特許庁手法」の妥当性を検証することであり、その中には、評価の定義及び評価サンプルの内容が、評価者に正しく理解されるものであるか、という観点も含まれる。このため、本調査においては、原則として各評価者の評価結果をそのまま採用する方針とし、フィードバック対象となった評価結果についても、これを修正するかどうかは評価者の判断に委ねた。

7.1.4. 人手評価補助ツール

「特許庁手法」では、その「事前準備」部分にて、「評価を簡便に実施するために必要なツール（人手翻訳補助ツール）」を作成し評価者に提供することが定められている（4.4.3.）。

本調査では、この「人手翻訳補助ツール」は、「品質評価対象データ」を収録した「品質評価対象データファイル」に組み込む形で実現した。詳しくは 6.3.3. 「品質評価作業用ファイルの作成」を参照のこと。

7.1.5. 評価実践にあたっての留意点

品質評価の実践にあたっては、「特許庁手法」に定められた要件に従い、特に以下の点に留意した。

7.1.5.1. 公平な評価

「特許庁手法」では、相対評価において評価対象となる各システムを公平に扱うため、「1つの機械翻訳対象文に対し複数の機械翻訳結果を評価する場合は、その全てを同一の評価者が同時に評価する」ことを要件としている。

本調査の品質評価実践においても、これを各評価者に周知し、また「品質評価対象データファイル」上でも各文に対する全ての機械翻訳結果を連続して表示させることで、この点に配慮した。

7.1.5.2. 「流暢さの評価」の取扱い

「流暢さの評価」は、原文の内容にとらわれず、機械翻訳結果の独立した文としての読みやすさ、理解し易さのみを評価するものである。「特許庁手法」では、この点に関して、『1つの機械翻訳対象文に対し同一の評価者が「流暢さ」の評価と他の観点からの評価の双方を担当する場合、最初に「流暢さの評価」を、原文や人手翻訳を参照せずに行う』ことを要件としている（4.3.3.4.）。

本調査はこの条件に該当するため、上記要件を評価者に周知するとともに、「品質評価対象データファイル」において「流暢さの評価」の評価結果入力欄を最も手前に配置し、これを無意識に行える環境を整備した。

7.2. 品質評価結果データ

7.1.にて実践した品質評価によって得られた各観点の評価結果を「品質評価対象データ」に反映し、「品質評価結果データ」を作成した。以下、その詳細を記す。

7.2.1. 品質評価結果の統合

評価者に配布する「品質評価対象データファイル」は、技術分野（IPC セクション）別に作成したが、「品質評価結果データ」は、全技術分野の評価結果を統合して1ファイル（1シート）とした。

また、「(1) 内容の伝達レベルの評価」は1文に対し3名の評価者による重複評価を行なったが、これらも全て1シートに統合した。

7.2.2. 分析用情報の追加

「品質評価結果データ」は、「相対評価」や「絶対評価」をはじめとする品質評価結果の分析に用いられる。このため、分析の際に必須となる以下の情報をあらかじめ組み込むこととした。

- ・各文における「内容の伝達レベルの評価」の3評価者評価値の平均値
- ・「内容の伝達レベルの評価」の文長（短中長）別評価値
- ・「重要技術用語の翻訳精度の評価」の評価値別カウント
- ・「流暢さの評価」の文長（短中長）別評価値

7.2.3. 「品質評価結果データ」のレイアウト

「品質評価結果データ」のレイアウトは、「作業用ファイル」のフォーマット（6.3.3.5.）をベースとして、これに7.2.1.～2.で説明した内容を反映させた。

次表 7-2 に、「品質評価結果データ」のフォーマットを示す。但し「作業用ファイル」と内容が同一の項目については、説明を省略する。

表 7-2. 「品質評価結果データ」のフォーマット

項 目	Excel 列	内 容
No. ～ メモ欄	A～L	※「作業用ファイル（表 6-6）」と同内容。
内容伝達結果①～③	M～O	「(1) 内容の伝達レベルの評価」の評価値（5～1）を、3 名の評価者別にセット。
内容伝達平均値	P	「内容伝達結果①～③」の平均値を文ごとに算出してセット。
内容伝達文長別平均値（短中長）	Q～S	「内容伝達平均値」を、原文の文長別（短文、中文、長文）に振り分けてセット。各列の集計値から文長別平均値が得られる。
重要技術用語①～③ 及び対訳	T～U, AA～AB, AH～AI	※「作業用ファイル（表 6-6）」と同内容。
重要技術用語①～③ 評価値	V, AC, AJ	「重要技術用語①～③」に対する「重要技術用語の翻訳精度」の評価値（A～D）をセット。
重要技術用語①～③ 評価値別カウント	W～Z, AD～AG, AK～AN	「重要技術用語①～③評価値」を、評価値別に振り分けて 1 をセット。
重要技術用語数	AO	※「作業用ファイル（表 6-6）」と同内容。
流暢さ評価値	AP	「流暢さの評価」の評価値をセット。
流暢さ文長別評価値 （短中長）	AQ～AS	「流暢さ評価値」を、原文の文長別に振り分けてセット。
【必須項目】A～G カ ウント結果	AT～AZ	「チェックリスト評価」の【必須項目】各項目のカウント値をセット。
A～G 合計	BA	各文の【必須項目】カウント合計値をセット。
【任意項目】及び【翻 訳言語別任意項目】カ ウント欄	BB～CC	「チェックリスト評価」の【任意項目】及び【翻訳言語別必須項目】の各項目のカウント値をセット。

なお、「品質評価結果データ」は、5 種類の機械翻訳結果のデータを全て収録したシートと、エンジン（RBMT1～SMT5）別のシートにて構成した。

「品質評価結果データ」のデータサンプルは、添付資料 7-7 を参照のこと。

第 8 章

品質評価結果の分析

第 8 章 品質評価結果の分析

「特許庁手法」を用いた品質評価の実践により取得した「品質評価結果データ」の内容を分析し、現状の英日機械翻訳に関する考察と、「特許庁手法」の品質評価手法としての妥当性の検証を行なった。本章にて、その内容を評価観点別にまとめる。

なお、本章における考察・検証内容には、第 2 回委員会¹における各委員からの指摘内容も反映してある。

8.1. 「品質評価結果データ」に関する注意事項

本調査にて取得した「品質評価結果データ」に関する注意事項を以下にまとめた。

8.1.1. 機械翻訳結果の作成条件

本調査における品質評価実践は、「特許庁手法」の妥当性の検証を主目的とした試行評価であり、評価対象とした 5 種類の英日機械翻訳エンジンの精密な性能比較を目的としたものではない。このため、各エンジンはほぼ初期状態で使用しており、特許文献翻訳用に最適化していない。また、一部エンジンのみ実用環境を用いる等、エンジン間の翻訳条件も一定ではない。

したがって、本調査にて作成した「品質評価結果データ」をそのまま各エンジンの性能比較に利用することは推奨しない。

8.1.2. データ件数

「特許庁手法」では、品質評価対象文を 200 文程度と定めた。本調査の品質評価実践も、この定めに従い、英日テストセットから第 6 章 6.1.項に示した構成にて抽出した 200 文も評価対象としたが、このうち 1 文について、原文中にスペルミスが生じており、テストセットの要件に反していることが判明した。このため「品質評価結果データ」は、この文を除外²した 199 文について作成した。

また、評価対象文の文長構成に関しても、6.1.に示した予定の構成から若干の変更が生じ、下表 8-1 のとおりとなった。

表 8-1. 評価対象文 文長別構成

	1～25 語	25～50 語	51 語～	合計
文 数	101 文	88 文	10 文	199 文
(予定文数)	(100 文)	(90 文)	(10 文)	(200 文)

¹ 第 2 回委員会の主な提言・指摘事項は、添付資料 1-5 を参照。

² 米国特許、H セクション、発明の詳細な説明からの採用文。

なお、「特許庁手法」では、評価結果の文長別平均値を実際の特許文献の構成比率（長文 50%、中文 40%、短文 10%）に補正して評価結果スコアを算出するため（6.1.3.及び 4.3.8.3. 参照）、「品質評価結果データ」を厳密に予定どおりの文長構成としなくても問題はない。

8.2. 評価結果の分析と考察：「内容の伝達レベルの評価」

本項にて、「内容の伝達レベルの評価」の評価結果を分析・考察する。

8.2.1. 評価基準

「特許庁手法」に定められた「内容の伝達レベルの評価」の評価基準を示す。

機械翻訳結果が、原文の内容をどの程度正確に伝達しているかを、人手翻訳の内容に照らして以下の 5 段階の評価基準で主観的に評価する。

- 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%)

本評価観点、1 文につき 3 名の評価者が重複評価を実施している（7.1.2.）。各エンジンの「評価スコア」は、「特許庁手法」の定めに則り、以下の手順にて算出した。

- ① 各文について 3 評価者の評価値の平均値を算出
- ② 全文について、長文、中文、短文ごとの平均値を算出
- ③ 上記各平均値を実際の特許文献の文長比率³に補正して「評価スコア」を算出

³ 長文・中文・短文それぞれの平均値を、「英日機械翻訳の現状分析」に用いたサンプル 20 文献における比率である長文 1：中文 4：短文 5 の文長比率（6.1.3.参照）に補正した（下記算式による）。

長文の平均スコア×0.1+中文平均スコア×0.4+短文平均スコア×0.5＝評価スコア

8.2.2. 評価スコア

前記手順にて算出した RBMT1～SMT5 の評価スコアを表 8-2 に示す。

表 8-2. 「内容の伝達レベル」評価スコア

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
スコア	2.39	2.79	3.01	3.15	2.76
相対順位	5	3	2	1	4

この評価スコア及び RBMT1～SMT5 の相対順位は、評価者間の意識合わせの基準としてあらかじめ作成していたリファレンス評価結果とほぼ合致した（下表 8-3 参照）。このことから、今回得られた評価スコアは、十分な妥当性を有していると考えられる。

表 8-3. 「内容の伝達レベル」評価スコアとリファレンス評価スコアの比較

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
内容伝達 スコア	2.39 5	2.79 3	3.01 2	3.15 1	2.76 4
リファレン ス評価	2.39 5	2.88 3	3.09 2	3.16 1	2.68 4
差異	0.00	+0.09	+0.08	+0.01	-0.08

8.2.3. 評価者別の評価値

「内容の伝達レベルの評価」は、1 文に対し 3 名の評価者による重複評価を実施している。この 3 名の評価者の評価結果を比較することで、「特許庁手法」の評価基準が、評価者が異なっても評価結果に大きなずれが生じないものであるかを検証できる。各評価者の評価結果が十分に整合していれば、「特許庁手法」の評価基準は、十分な客観性を有する評価結果を得られる妥当な基準であるといえる。

下表 8-4 に、3 名の評価者（A～C）別の評価スコアの比較結果を示す。

表 8-4. 「内容の伝達レベル」評価者別平均評価値

		RBMT1	RBMT2	RBMT3	SMT4	SMT5
A	スコア	2.28	2.74	2.92	3.05	2.64
	相対順位	5	3	2	1	4
B	スコア	2.87	3.31	3.46	3.54	3.21
	相対順位	5	3	2	1	4
C	スコア	2.01	2.32	2.64	2.86	2.43
	相対順位	5	4	2	1	3

3 評価者の評価結果を比較すると、評価者 C のみ、RBMT2 と SMT5 の順位が他の評価者と逆になっているものの、それ以外の各エンジンの相対順位は 3 者間で合致しており、少なくとも相対順位に関しては、3 者間で十分に整合していることが確認できた。したがって、本評価基準は、評価者による相対順位のばらつきが生じにくいものであるということが検証された。

一方、評価の水準（厳しさ、甘さ）については、各評価者間でやや開きがあり、評価値の最も高い評価者 B と最も低い評価者 C とでは、平均 0.82 点の差がある。こうした評価水準のずれは、相対評価においては大きな問題とはならないが、「内容の伝達レベルの評価」の評価スコアは絶対評価の基準にも用いられるものであり、正確な絶対評価を行うためには、評価水準が評価者間で揃うようにする必要がある。

今回、評価者間で水準のずれが生じた大きな理由として、事前の評価者間の意識合わせが不十分であった可能性が考えられる。

今回の品質評価実践では、評価者間の意識合わせは、作業期間等の都合上、各評価者の 1 日分の評価結果を個別に回収し、あらかじめ準備したリファレンス用の評価結果と比較して評価値が 2 以上違うものに対して、リファレンス評価の評価理由を添えて再確認を促す形とした（7.1.3.2.参照）。したがって、各評価者間の水準のずれという観点からは、最大 2（-1～+1）までは不問とされ得る。

今回の結果から、上記手法による意識合わせでは評価者間の水準を十分に合わせる事が難しいことが判明したため、より適切な手法に改善する必要がある。この点については、「特許庁手法」の課題のひとつとして、第 9 章 9.1.1.にて再度考察する。

ただし、今回の評価結果に関しては、評価者 A の評価水準がリファレンス評価結果とほぼ同等、評価者 B がやや高く、評価者 C がやや低いという分布となっており（表 8-3～8-4 参照）、3 評価者を平均した評価スコアは、結果的にはリファレンス評価にかなり近い水準となった。したがって、今回得られた評価スコア自体は、想定していた評価水準を大きく逸脱するものではなく、後述の絶対評価の基準の検証（8.6.項）に利用できる程度の妥当性は有しているものと考ええる。

8.2.4. 符号検定による検証

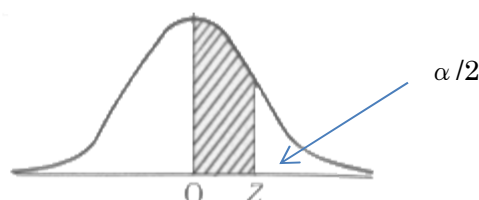
「内容の伝達レベル」の評価結果に基づく相対評価について、各エンジンの評価スコア間にどの程度の差があれば、その差異に基づく相対順位に十分な有意性があるかを検証した。検証は「符号検定」にて行い、8.2.2.に示した評価スコアに基づく各エンジン（RBMT1～SMT5）の相対順位の有意差を検定した。

符号検定は以下のように行われる⁴。

エンジン i とエンジン j の翻訳精度評価結果を比較して i が j より高い評価を得たデータ数を n1、低い評価を得たデータ数を n2 とし、 $n=n1+n2$ とする。この時、

$$z = \frac{|n1-n2|-1.0}{\sqrt{n}}$$

を求め、標準正規分布の z 以上の部分の確率を $\alpha/2$ とする(下図参照)。そのとき $1-\alpha$ (信頼水準) はエンジン i とエンジン j の精度の差が有意である確率を示す。



※グラフは http://www.koka.ac.jp/morigiwa/sjs/standard_normal_distribution.htm より引用

以下、符号検定結果を示す。

表 8-5. 符号検定による「内容の伝達レベル」評価スコア比較

A vs B	A 上位(n1)	同点	B 上位(n2)	Z	1.0- α (信頼水準)	差異※
RBMT1 vs RBMT2	47	34	118	5.45	100.0%	-0.40
RBMT1 vs RBMT3	45	20	134	6.58	100.0%	-0.62
RBMT1 vs SMT4	43	21	135	6.82	100.0%	-0.76
RBMT1 vs SMT5	64	26	109	3.35	99.9%	-0.37
RBMT2 vs RBMT3	66	32	101	2.63	99.1%	-0.22
RBMT2 vs SMT4	60	28	111	3.82	100.0%	-0.36
RBMT2 vs SMT5	89	18	92	0.15	11.8%	-0.03
RBMT3 vs SMT4	76	25	98	1.59	88.9%	-0.14
RBMT3 vs SMT5	108	19	72	2.61	99.1%	0.25
SMT4 vs SMT5	109	43	47	4.88	100.0%	0.39

※比較した MT (A と B) の内容伝達スコアの差異 (A-B)。

⁴ たとえば以下を参照： http://arena.hyogo-dai.ac.jp/~kawano/HStat/?2009%2F13th%2FSign_Test

表 8-5 のとおり、「内容の伝達レベルの評価」のスコア比較により得られた各エンジン間の相対順位は、その大半において信頼水準 ($1.0 - \alpha$) が 99% 以上であり、高い有意性が検定された。ただし、信頼水準は一般に 95% 以上で有意性ありと見なされることを考えると、RBMT2 と SMT5 (内容伝達スコアの差異-0.03) の相対順位は信頼水準が極めて低く、また RBMT3 と SMT4 (差異-0.14) についても有意性は充分ではない。

これらの結果から、「特許庁手法」の内容伝達スコアによる相対順位比較は、評価対象文が 200 文程度である場合、両者のスコア差異が 0.20 以上であれば、十分な信頼性を有すると判断できる。

同時に、このことは、RBMT2 と SMT5 のように性能が接近しており、内容伝達スコアの差異が 0.20 に満たないエンジンに対して「内容の伝達レベル」の観点から優劣を決める必要がある場合には、評価対象文数は 200 文では不十分であり、より多くの文数が必要であるということも意味している。8.2.3.にて評価者 C のみ RBMT2 と SMT5 の順位が入れ替わっていたのも、評価基準自体の問題ではなく、この理由によるものと考えられる。

ただし、「特許庁手法」には「内容の伝達レベル」の他にもいくつかの品質評価観点が設けられており、また、エンジンの総合的な優劣評価は、翻訳品質以外にも、例えば価格やチューニングの容易さ等、いくつかの検討項目があるため、必ずしも 1 つの評価観点において常に優劣をつける必要はない。したがって、「特許庁手法」の設定する 200 文という評価対象文数は、上記結果からは、妥当なデータ量であるといえる。

※注意

符号検定は母集団からのランダムサンプルを仮定している。これに対して今回の検定の標本は、実際の特許文献の文長構成よりも長文の比率を低く設定しており、厳密にはランダムではないが、下記のとおりその影響は軽微と考える。

- ・ 実際の特許文献の文長比率＝長文：1、中文：4、短文：5
- ・ 標本 199 文の文長比率＝長文：0.5、中文：4.4、短文：5.1

8.2.5. 評価対象文数の妥当性の検証

「特許庁手法」では、「内容の伝達レベル」（及び「重要技術用語の翻訳精度」）の評価対象件数を 200 文と定めたが、この文数にて正確な品質評価値が得られるかどうかを、今回の評価対象文を半数ずつに分け、双方の評価結果が整合するかによって検証した。

今回の評価対象文 199 文を、IPC でソートしたうえ交互に 2 組に振り分け、100 文の組（1 組）と 99 文の組（2 組）とし、それぞれについて「内容伝達レベル」の評価スコアを比較した。結果を次表に示す。

表 8-6. 「内容の伝達レベル」100 文ずつの評価スコア

		RBMT1	RBMT2	RBMT3	SMT4	SMT5
1 組	平均値	2.33	2.77	3.01	3.17	2.71
	相対順位	5	3	2	1	4
2 組	平均値	2.44	2.80	3.00	3.13	2.80
	相対順位	5	3	2	1	4
平 均	平均値	2.39	2.79	3.01	3.15	2.76
	相対順位	5	3	2	1	4

比較の結果、1 組と 2 組の評価スコアは、IPC がほぼ同数となること以外、文長や記載箇所等は無作為に組分けしたにも関わらず、相対順位、スコア値とも、ほぼ均一の結果となった。この結果から、「特許庁手法」の定める評価対象文数 200 文は、正確な品質評価を行うのに十分なデータ量であると判断できる。

ただし、8.2.4.に示した符号検定による検証結果のとおり、比較するエンジンの性能が非常に接近している場合、正確な相対順位を得るためには評価文数は 200 文では不十分な場合もあるので留意が必要である。

8.3. 評価結果の分析と考察：「重要技術用語の翻訳精度の評価」

本項では「重要技術用語の翻訳精度の評価」の評価結果についての分析と考察を行う。

8.3.1. 評価基準

「特許庁手法」の「重要技術用語の翻訳精度の評価」の評価基準を以下に示す。

A（適訳）：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）：技術用語として一般的に用いられる訳語ではないが、意味はおおむね正しい。
C（誤訳）：誤訳である。
D（不訳）：未知語、訳漏れである。
※文中に同一の重要技術用語が複数回出現し、異なる評価値となる場合は、最も良い評価値が採用される。

本調査の品質評価実践では、評価対象文 199 文の各文においてあらかじめ選定されている重要技術用語（1 文あたり最大 3 語）それぞれに対し、上記基準にて翻訳精度の評価を行い、評価結果を集計して、199 文における重要技術用語の総語数（524 語）に対する各評点（A～D）の比率を算出した。

8.3.2. 評価スコア

下表 8-7 に、RBMT1～SMT5 それぞれについて、評点 A～D 別の語数及び比率を示す。

表 8-7. 「重要技術用語の翻訳精度」評価結果

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
評価 A (適訳語)	273 語 52.1%	317 語 60.5%	348 語 66.4%	443 語 84.5%	394 語 75.2%
評価 B (可訳語)	108 語 20.6%	87 語 16.6%	90 語 17.2%	41 語 7.8%	46 語 8.8%
評価 C (誤訳語)	136 語 26.0%	110 語 21.0%	83 語 15.8%	39 語 7.4%	61 語 11.6%
評価 D (不訳語)	7 語 1.3%	10 語 1.9%	3 語 0.6%	1 語 0.2%	23 語 4.4%
相対順位 (適訳率)	5	4	3	1	2

8.3.3. 符号検定による検証

「重要技術用語の翻訳精度」の評価結果についても、「内容の伝達レベル」と同様、符号検定による検証を行なった。検定には、評価対象文全文にて選定された全ての重要技術用語（524 語）における評価値の差異を用いた。下表 8-8 に、その結果を示す。

表 8-8. 符号検定による「重要技術用語の翻訳精度」適訳率比較

A vs B	A 上位(n1)	同点	B 上位(n2)	Z	1.0－ α (信頼水準)	差異※
RBMT1 vs RBMT2	83	310	131	3.21	99.9%	-8.4%
RBMT1 vs RBMT3	56	322	146	6.26	100.0%	-14.3%
RBMT1 vs SMT4	31	282	211	11.51	100.0%	-32.4%
RBMT1 vs SMT5	69	276	179	6.92	100.0%	-23.1%
RBMT2 vs RBMT3	65	352	107	3.13	99.8%	-5.9%
RBMT2 vs SMT4	38	311	175	9.32	100.0%	-24.0%
RBMT2 vs SMT5	75	307	142	4.48	100.0%	-14.7%
RBMT3 vs SMT4	43	342	139	7.04	100.0%	-18.1%
RBMT3 vs SMT5	81	328	115	2.36	98.2%	-8.8%
SMT4 vs SMT5	93	396	35	5.04	100.0%	9.3%

※比較した MT（A と B）の適訳率の差異（A－B）。

上表のとおり、全ての相対順位において信頼水準（1.0－ α ）は 98%以上であり、評価値に基づく相対順位（MT4>MT5>MT3>MT2>MT1）の高い有意性が検定された。この結果から、重要技術用語の翻訳精度の評価という観点において、524 語という評価対象語数は相対評価には十分なデータ数であり、したがって 200 文、1 文あたり最大 3 語という「特許庁手法」の設定は妥当である、と判断できる。

ただし、上記結論は、性能が極めて接近しているエンジン同士の比較においては必ずしもその限りではない。今回の検定結果からは、例えば適訳率において、少なくとも 6%程度の相違があれば、相対順位の有意性は高いといえる。

8.4. 評価結果の分析と考察：「流暢さの評価」

「流暢さの評価」の評価結果と、その分析・考察内容について本項にまとめた。

8.4.1. 評価基準

「特許庁手法」の「流暢さの評価」の評価基準を示す。

機械翻訳結果を独立した文として見た際の、文としての自然さ、読みやすさのみを以下の 5 段階の評価基準で主観的に評価する。

- 5：文意が明解で、人間が書いた文に近い。
- 4：文として不自然な箇所を若干含むが、文意は明解である。
- 3：文として不自然な箇所があり、文意がわかりにくい。
- 2：文法規則に反する表現をかなり含む。文意がわからない。
- 1：文として成立していない。

＜機械翻訳対象文のうちランダムに選択した 50 文のみを対象とするサンプリング評価を行う。一定数の長文を含める＞

なお、「特許庁手法」は上記のとおり「流暢さの評価」を 50 文のみのサンプリング評価としているが、今回の品質評価実践においては、このサンプリング文数の妥当性を検証する必要もあることから、機械翻訳対象文の全文（199 文）を評価対象とした。

8.4.2. 評価スコア

表 8-9 に、「流暢さの評価」の評価スコアを示す。なお、この評価スコアは、「特許庁手法」の定めに従い、文長別（短・中・長）に算出した平均評価値を、実際の特許文献の文長比率（5:4:1）に合わせて補正した数値⁵とした。

表 8-9. 「流暢さ」評価スコア

	MT1	MT2	MT3	MT4	MT5
スコア	2.46	2.78	3.02	3.16	2.83
相対順位	5	4	2	1	3

エンジン間の相対順位は、「内容の伝達レベル」の評価スコア、及び「重要技術用語の翻訳精度」の適訳率とほぼ合致するものの、RBMT2 と SMT5 が僅差ながら逆転した。

⁵ 8.2.1.の脚注 3 参照。

8.4.3. 符号検定による検証

「流暢さの評価」の相対順位の有意性についても、符号検定による検証を行なった。下表 8-10 にその結果を示す。

表 8-10. 符号検定による「流暢さの評価」評価値比較

A vs B	A 上位(n1)	同点	B 上位(n2)	Z	1.0- α (信頼水準)	差異※
RBMT1 vs RBMT2	39	71	89	4.33	100.0%	-0.33
RBMT1 vs RBMT3	33	65	101	5.79	100.0%	-0.57
RBMT1 vs SMT4	31	51	117	6.99	100.0%	-0.75
RBMT1 vs SMT5	49	59	91	3.47	99.9%	-0.39
RBMT2 vs RBMT3	43	86	70	2.45	98.6%	-0.24
RBMT2 vs SMT4	47	64	88	3.44	99.9%	-0.42
RBMT2 vs SMT5	65	63	71	0.43	33.2%	-0.05
RBMT3 vs SMT4	54	67	78	2.00	95.5%	-0.18
RBMT3 vs SMT5	77	67	55	1.83	93.2%	0.18
SMT4 vs SMT5	84	78	37	4.18	100.0%	0.36

※比較した MT (A と B) の内容伝達スコアの差異 (A-B)。

他の評価スコアと同様、「流暢さの評価」においても、大半の相対順位は信頼水準 (1.0 - α) が 99%前後であり、高い有意性が検定された。唯一、RBMT2 と RBMT5 (適訳率の差異-0.05) の相対順位のみ信頼水準が極めて低く、また、RBMT3 と SMT5 (差異 0.18) においても、信頼水準が、有意性ありとみなされる 95%を若干下回る。

これらの結果から、「流暢さの評価」の評価スコアによる相対順位比較は、両者の適訳率等の差異が約 0.20 以上の場合に、十分な信頼性を有すると判断できる。

8.4.4. 評価対象文数の妥当性の検証

「特許庁手法」では、「流暢さの評価」の対象文数をサンプリングによる 50 文と定めている。この評価文数の妥当性について検証するため、品質評価実践では、「流暢さの評価」を 199 文全文について実施し、199 文全文の平均値と、199 文をランダムに 4 組（49 文～50 文）に分けたそれぞれの平均値とを比較して、近似の結論が得られるか否かを調査した。

表 8-11 に、199 文全体の「流暢さの評価」平均評価値および 4 組それぞれの平均評価値を示す（※いずれも文長別補正前）。

表 8-11. 「流暢さの評価」50 文サンプリング比較結果

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
199 文平均	2.48 5	2.81 4	3.05 2	3.23 1	2.86 3
第 1 組	2.44 5	2.70 4	3.06 2	3.26 1	2.76 3
第 2 組	2.36 5	2.88 3	2.96 2	3.06 1	2.82 4
第 3 組	2.56 5	2.82 4	3.14 2	3.28 1	2.88 3
第 4 組	2.55 5	2.84 4	3.02 2	3.31 1	3.00 3

上表のとおり、第 2 組にて RBMT2 と SMT5 の順位が逆転している以外は、いずれの組においても 199 文と同一の相対順位が得られた。この結果からは、相対評価の補助的項目として使用する分には、「流暢さの評価」は 50 文のサンプル評価でも充分なように見える。

しかしながら、各組にて符号検定を行なってみると、下表 8-12 のとおり信頼水準が 95% に達しない組み合わせは多い。これは、厳密な評価を行うためには 50 文では不足であることを意味し、他の評価と同様、全文（200 文）を評価対象とすることが望ましいといえる。ただし、「流暢さの評価」は重要度が比較的低いため、評価リソースとの兼ね合いとなる。

表 8-12. 符号検定による「流暢さの評価」サンプル評価値比較（信頼水準%）

	1vs2	1vs3	1vs4	1vs5	2vs3	2vs4	2vs5	3vs4	3vs5	4vs5
第 1 組	95.5	99.8	100	87.7	90.0	99.1	28.1	77.8	75.7	99.3
第 2 組	99.6	100	99.7	95.1	55.9	40.4	0.0	15.0	52.8	89.0
第 3 組	96.3	97.4	100	39.3	90.0	75.7	0.0	42.9	90.0	98.2
第 4 組	77.0	96.9	99.9	98.2	15.5	95.5	48.9	83.6	0.0	71.9

8.5. 評価結果の分析と考察：「チェックリストに基づく評価」

今回の品質評価実践における「チェックリストに基づく評価」の評価結果に関する分析・考察の内容を本項に記す。

8.5.1. 評価基準

「チェックリストに基づく評価」は、機械翻訳結果中の全ての翻訳ミスを1つずつ分類しカウントしていく方式の評価項目である。「特許庁手法」においては、評価対象となるシステムの翻訳ミスの傾向や弱点を把握し、性能改善のためのフィードバックを行うために実施する評価として扱われており、相対評価、絶対評価には用いられない。

チェックリストに含めるチェック項目として、言語共通の【必須項目】7項目、同じく言語共通の【任意項目】12項目、そして翻訳言語に応じた【翻訳言語別任意項目】として英日16項目が設定されている。ただし、このうち【任意項目】と【翻訳言語別任意項目】は、評価の目的等に合わせて任意に選択可能である。また、評価対象文数も任意設定とされている。

本調査の品質評価実践では、各チェック項目が正しく機能するかを検証する必要があるため、【任意項目】、【翻訳言語別任意項目】を含め、全てのチェック項目について評価した。また、評価対象文数も199文全文を対象とした。

8.5.2. 【必須項目】カウント集計結果

表8-13に、「チェックリストに基づく評価」の【必須項目】7項目の項目別カウント集計結果をエンジン別に示す。

表8-13. 「チェックリスト評価」【必須項目】カウント集計結果

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
A 語の欠落	18	13	18	32	100
B 不要語の混入	11	4	11	18	26
C 未知語	9	11	11	5	30
D 品詞誤り	83	59	46	55	55
E 係り受け誤り	164	127	129	140	158
F 語の誤訳	420	355	270	119	162
G その他	30	14	20	42	29
合 計	735	583	505	411	560
(少ない順)	5	4	2	1	3

「チェックリストに基づく評価」は、前述のとおり、評価対象となるシステムの翻訳ミスの傾向や弱点を把握する目的で行われるものである。したがって、この項目の妥当性の検証は、その評価結果から各システムの弱点や翻訳ミスの傾向が読み取れるか否かにより行うべきである。

これを踏まえて上記カウント集計結果を見ると、例えば **SMT5** では「A. 語の欠落」が他のエンジンに比べて非常に多いことや、**RBMT1** では品詞誤りが他のエンジンに比べて多めであること、**SMT** は **RBMT** に比べて「A 語の欠落」や「B 不要語の混入」が起りやすく、一方 **RBMT** では語の誤訳が起りやすいこと、等の傾向を読み取ることができる。

ただし、これらは全て複数のエンジンの集計結果の比較によって判明したことである。「チェックリストに基づく評価」は必ずしも複数のエンジンを対象に行うとは限らず、単一のシステムを評価する場合は、他との比較による傾向の把握は難しくなる。また、対象とする文や文数は評価の都度変化することが想定されるため、どの程度のカウント数が平均値であるかを定めにくく、したがって目安となるカウント数をあらかじめ設定しておくことも難しい。

これらを考えると、「チェックリストに基づく評価」は、仮に評価対象とするシステムが1種のみであったとしても、比較用として複数のシステムを同時に評価することが推奨される。

なお、「チェックリストに基づく評価」の評価結果は、相対評価や絶対評価には用いない方針であるが、こと【必須項目】に限れば、全ての翻訳ミスが必ずカウントされるという性質上、他の評価観点のスコアとの一致度は比較的高く、各エンジンの相対順位（翻訳ミスの少ない順）は、「内容の伝達レベル」や「重要技術用語の翻訳精度」の評価結果とほぼ同じ結果となっている。

8.5.3.【任意項目】カウント集計結果

続いて、RBMT1～SMT5の【任意項目】別カウント集計結果を示す。

表 8-14.「チェックリスト評価」【任意項目】

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
Z00 訳語ゆれ	15	11	16	13	15
Z01 文の分断	10	14	5	27	3
Z02 主語述語	44	28	18	37	40
Z03 並列表現	59	54	43	53	54
Z04 名詞句	5	2	0	1	3
Z05 特許特有用語	63	54	41	27	44
Z06 特許特有記号	14	9	6	13	6
Z07 箇条書き	13	4	10	28	27
Z08 カッコ書き	6	4	7	5	6
Z09 数式化学式	5	3	4	4	3
Z10 数字と単位	6	3	2	1	1
Z11 数値範囲	8	9	9	5	8
合 計	248	195	161	214	210
	5	2	1	4	3

【任意項目】は、機械翻訳全般において特に課題となる特定のパターンの翻訳ミスのみをカウントする項目である。例えば「Z-08 カッコ書きに関する翻訳ミス」や「Z-09 数式化学式に関する翻訳ミス」など、評価対象となる文（カッコ書き／数式化学式を含んでいる文）自体が少数であるものも多いが、上記のとおり、全ての項目においてカウントが生じており、設定された各チェック項目が妥当なものであることを示している。

各エンジンの傾向としても、SMT4は「Z-01 文の分断」に該当する翻訳ミスが多く、また SMT4 及び SMT5 で「Z-07 箇条書き」に該当する翻訳ミスも比較的多いことがわかる等、ある程度読み取ることができる。

【任意項目】の評価結果を利用したフィードバックの方法としては、対象とする機械翻訳エンジンが特に苦手とする課題や、機械翻訳結果の実用面から特に優先的に改善したい課題について、該当するチェック項目にカウントされた文のみを抽出することで、効率的かつ具体的なフィードバックを行うことが考えられる。

なお、【任意項目】における相対順位（カウントの少なさ）は、他の評価観点や【必須項目】とは全く異なる結果となったが、これは、【任意項目】が翻訳ミスのうち特定のケースのみをカウントするものである以上、十分に起こり得るものである。

8.5.4.【翻訳言語別任意項目 [英日]】カウント集計結果

最後に、【翻訳言語別任意項目 [英日]】のカウント集計結果を示す。

表 8-15. 「チェックリスト評価」【翻訳言語別任意項目 [英日]】

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
EJ01 ピリオド	0	2	0	5	0
EJ02 コロン類	4	0	1	2	5
EJ03 重要単複情報	4	4	4	3	3
EJ04 単複併記	5	3	5	4	5
EJ05 序数詞表現	9	5	2	2	1
EJ06 語の挿入	7	6	7	5	7
EJ07 倒置文	5	6	4	6	6
EJ08 名詞代名詞	2	1	2	1	1
EJ09 慣用表現	33	30	22	17	21
EJ10 前置詞	47	31	38	27	36
EJ11 接頭接尾辞	8	8	2	2	5
EJ12 スラッシュ	3	7	1	3	5
EJ13 therein 等	4	2	9	7	15
EJ14 代名詞 same	5	5	5	1	1
EJ15 小文字混入	0	1	0	0	0
EJ16 クレーム引用	12	9	13	10	13
合 計	151	124	117	111	137
	5	3	2	1	4

【翻訳言語別任意項目 [英日]】に関しては、【任意項目】以上に対象文数が限定される場合が多く、このため相対比較でも目立つ数字は少ないが、それでも全項目で翻訳ミスがカウントされており、項目設定が妥当であることを示している。

8.6. 絶対評価スコアのマッチング基準の妥当性検証

「特許庁手法」では、絶対評価（品質評価値から各種実務用途への有用性を直接判断）の方法として、各種用途と「内容の伝達レベル」の評価スコアとを、以下の基準でマッチングさせることとしている（詳細は第4章4.3.9.参照）。

3. 絶対評価方法

(0) 絶対評価スコアと用途とのマッチングは、以下のとおりとする。

スコア5：権利認定、引用発明認定に適した翻訳品質

スコア4：実体審査、先行技術調査における精読に適した翻訳品質

スコア3：先行技術調査及び文献紹介時の粗読及び検索データに適した翻訳品質

スコア2：粗読又は検索データに適した翻訳品質

スコア1：特許調査に際しては全く利用できない翻訳品質

(1) 「内容の伝達レベルの評価」のスコアに基づき、絶対評価スコアを定める。

内容伝達スコア：4.5～5.0 → 絶対評価スコア：5

内容伝達スコア：3.5～4.5 → 絶対評価スコア：4

内容伝達スコア：2.5～3.5 → 絶対評価スコア：3 ← マッチング基準

内容伝達スコア：1.5～2.5 → 絶対評価スコア：2

内容伝達スコア：1.0～1.5 → 絶対評価スコア：1

(2) 「重要技術用語の翻訳精度の評価」の「適訳率」が一定値未満の場合は、絶対評価スコアの上限を2とし、これを超えている場合はスコアを2に下げる。一方、「適訳率」が一定値以上の場合は、絶対評価スコアの下限を2とし、これに満たない場合はスコアを2に上げる。

(3) 「流暢さの評価」において、特筆すべき問題点が存在する場合は、その内容に応じてスコアを減点する。

ただし、上記(1)に示したマッチング基準は、妥当性未検証のまま「仮設手法」のものをそのまま採用しており（4.3.9.5.）、今回の品質評価実践の結果を用いて、その妥当性を検証する必要がある。以下、その検証結果について示す。

8.6.1. 妥当性検証の方法

マッチング基準の妥当性検証は、元特許庁審査官（1名）を評価者に起用し、199文の機械翻訳結果それぞれが、前記(0)に定義された絶対評価スコア 5～1 のどれに該当するかを、文単位で直接スコア付けする方式で実施した。

なお、評価するのはあくまで翻訳の品質であり、原文の内容の技術的重要度は評価値に反映させるべきではない。したがって、絶対評価スコアの付与は、「評価対象文の機械翻訳結果の翻訳精度で文献全体が翻訳されていると仮定した時、その文献はどのレベルの特許実務に有用であると思われるか」という観点で行なった。

8.6.2. 集計結果

199文に対して付与された絶対評価スコアを文長別に集計し、実際の特許文献の文長構成に基づく補正⁶を施した結果を以下に示す。以後、これを「検証用スコア」と呼ぶ。

表 8-16. 「絶対評価スコア」平均値（＝検証用スコア）

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
検証用スコア	1.68	2.12	2.40	2.82	2.37
相対順位	5	4	2	1	3

検証用スコアの相対順位は、「内容の伝達レベル」の評価とほぼ合致しており、「内容の伝達レベル」を絶対評価の基準とする現行の手法は妥当であると考えられる⁷。RBMT2 と SMT5 のみ順位が逆転しているが、これは「重要技術用語の翻訳精度」の違い（適訳率＝RBMT2：60.5%、SMT5：75.2%）が影響したものと思われ、この評価観点をマッチング基準に反映させるという方針も妥当といえる。

ただし、この検証用スコアを四捨五入した値と、「内容の伝達レベル」評価値と現行マッチング基準にて算出した絶対評価スコアとを比較すると、表 8-17 のとおり、両者が整合しないケースが見られる。

表 8-17. 「絶対評価スコア」平均値

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
検証用スコア	1.68	2.12	2.40	2.82	2.37
四捨五入	2	2	2	3	2
内容伝達 現行スコア	2.39 2	2.79 3	3.01 3	3.15 3	2.76 3

⁶ 8.2.1.の脚注 3 参照。

⁷ 検証用スコアと内容伝達スコアの Pearson の積率相関係数は 0.762（かなり強い相関がある）と算出された。

具体的には、3つのエンジン（RBMT2、RBMT3、SMT5）において、検証用スコアの四捨五入値が2であるのに対し、現行マッチング基準により算出された絶対評価スコアが3となっており、両者が整合しない。したがって、現行マッチング基準には改善の余地があるといえる。

8.6.3. マッチング基準の見直しの観点

現行マッチング基準の見直しにあたり、まずは絶対評価スコア1～3の関係について再考する。8.6.1.に示したとおり、現行のマッチング基準では、絶対評価スコア1～3に、それぞれ異なる内容伝達スコアの範囲（1.0～1.5、1.5～2.5、2.5～3.5）が対応づけられている。

しかしながら、これら1～3の絶対評価スコアで考慮されるのは「粗読」と「検索」の2つの用途に有用であるか否かのみであり、主に「粗読」＝内容伝達スコア、「検索」＝技術用語適訳率でそれぞれの適否を評価するとすれば、下記のとおり、それぞれについて1つずつ閾値を設ければ充分といえる。

絶対スコア	粗読（内容伝達スコア）	検索用途（技術用語適訳率）
スコア1	閾値未満	閾値未満
スコア2	いずれか一方のみ閾値以上	
スコア3	閾値以上	閾値以上

なお、絶対評価スコア4～5についても、「特許庁手法」における定義では「適訳率」が一定値（＝閾値）未満の場合はスコアを2に下げることになっており⁸、したがってマッチング基準は、以下のようなマトリックスで示すのが適当である。

内容伝達 適訳率	1.0～ n.n	n.n ～3.5	3.5～4.5	4.5～5.0
nn %以上	2	3	4	5
nn %未満	1	2		

以上を踏まえ、上記の新マッチング基準において絶対評価スコア1～3に対応する内容伝達スコア（**n.n**）及び技術用語適訳率（**nn**%）の閾値を検討する。

⁸ 8.6.1.に示した「特許庁手法」の「絶対評価方法」(2)を参照。

8.6.4. 内容伝達スコアの閾値の検討

内容伝達スコアの閾値の検討として、元審査官によって付与された文単位の絶対評価スコア（以下「審査官絶対スコア」という）と、各スコアに該当する文の「内容伝達レベル」の平均値との関係を、全文（199 文×5MT）を対象に調査した。以下に結果を示す。

表 8-18. 「絶対評価スコア」平均値

審査官絶対スコア	5	4	3	2	1
文数	53	145	214	215	368
内容伝達平均値	4.51	3.83	3.28	2.68	1.98
(現行スコア)	(5)	(4)	(3)	(3)	(2)

このうち審査官絶対スコア 4～5 に関しては、現行マッチング基準と正しく対応しており、それぞれのマッチング基準（3.5～4.5、4.5～5.0）は妥当といえる。したがって、スコア 1～3 について、「粗読」に最低限必要なスコア（閾値）を、1.0～3.5 の範囲内で定めればよいこととなる。

上表の結果からは、内容伝達スコアが 1.98（＝検証用スコア 1）では「粗読」の役に立たず、3.28（＝検証用スコア 3）ならば役に立つと判断できる。したがって、閾値は 2.0～3.0 の範囲内に定めるのが妥当と考えられる。

8.6.5. 技術用語適訳率の閾値の検討

「検索」用途への適否を測る技術用語適訳率の閾値の検討も、8.6.4.と同様、審査官絶対スコア別の「適訳率」の平均値を、全文（199 文×5MT）にて調査することで行なった。下表 8-19 に集計結果を示す。

表 8-19. 「技術用語適訳率」平均値

審査官絶対スコア	5	4	3	2	1
文数	53	145	214	215	363
適訳率平均値	92.8%	79.0%	71.3%	66.2%	57.1%

上記結果からは、適訳率が 57.1%以下（＝検証用スコア 1）のものは「検索」用途には不足であり、71.3%以上（＝検証用スコア 3）であれば有用と判断できる。したがって、「適訳率」の閾値は、60%～70%の範囲内と考えられる。

8.6.6. 新マッチング基準とその検証

上記より、「粗読」用途に有用となる内容伝達レベルの閾値を 2.0～3.0 の範囲、「検索」用途に有用となる適訳率の閾値を 60%～70%の範囲に絞り込んだ。この範囲内で、検証用スコアとの齟齬を極力抑えるように数値を調整すると、それぞれ以下のとおりとなる。

(1) 絶対評価スコアを、「内容の伝達レベルの評価」のスコア及び「重要技術用語の翻訳精度の評価」の適訳率に基づき定める。

内容伝達 適訳率	1.0～ 3.0	3.0 ～3.5	3.5～4.5	4.5～5.0
60% 以上	2	3	4	5
60% 未満	1	2		

~~(2) 「重要技術用語の翻訳精度の評価」の「適訳率」が一定値未満の場合は、絶対評価スコアの上限を 2 とし、これを超えている場合はスコアを 2 に下げる。一方、「適訳率」が一定値以上の場合は、絶対評価スコアの下限を 2 とし、これに満たない場合はスコアを 2 に上げる。~~

※重要技術用語の「適訳率」は（１）のマトリックス内に反映されたため、（２）の定義は不要となり削除した。

上記新マッチング基準を具体的に表現すると、以下となる。

- ・平均して原文の内容の 50%以上が正しく伝達されていれば「粗読」に有用である
- ・文献中に出現する技術用語の 60%以上が通常同義語の範囲に訳されていれば「検索」に有用である

この新マッチング基準にて RBMT1～SMT5 の絶対評価スコアを改めて算出すると、以下のとおりの結果となる。

表 8-20. 「絶対評価スコア」平均値

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
伝達スコア	2.39	2.79	3.01	3.15	2.76
用語適訳率	52.1%	60.5%	66.4%	84.5%	75.2%
絶対評価スコア	1	2	3	3	2
検証用スコア	1.68 (2)	2.12 (2)	2.40 (2)	2.82 (3)	2.37 (2)

表 8-20 に示されるとおり、新マッチング基準によって算出した各エンジンの絶対評価スコアも、RBMT1 と RBMT3 にて、検証用スコアの四捨五入値と合致していない。だが、RBMT3 に関しては、伝達スコアが 3.01 とほぼ閾値上にあり、このエンジンに付与された絶対評価スコア 3 は、極めて 2 に近い 3 であるといえる。また、RBMT1 に関しては、内容の伝達レベル、用語適訳率とも他のエンジンとかなりの差があり、新マッチング基準に基づく絶対評価スコア（スコア 1）で妥当であると判断した。

ただし、今回設定した新マッチング基準は、あくまで 1 名の元審査官による 1 つの検証用スコアのみを根拠に定めたものであり、その妥当性についてはさらに検討・検証のうえ、基準を調整していく必要がある。

8.7. 「特許庁手法」の妥当性

今回の品質評価実践により、「特許庁手法」の妥当性に関し、以下の知見を得ることができた。

8.7.1. 評価対象文数

「内容の伝達レベルの評価」及び「重要技術用語の翻訳精度の評価」に関しては、「特許庁手法」に定めた 200 文程度を対象文とすることにより、十分な有意性をもつ相対評価結果が得られることが検証された。

ただし、性能がごく接近したエンジン同士の相対順位、具体的には内容伝達スコアの差異が 0.20 未満の範囲に入るエンジン間の相対順位は、200 文の評価では十分な有意性が得られないこともわかった。このような場合は、品質は同等と扱うべきであろう。

また、「流暢さの評価」については、50 文のみを対象としたサンプリング評価と定めているが、今回の検証では、50 文では信頼性の高い相対順位を得ることが難しいことがわかった。

8.7.2. 評価観点・評価基準の妥当性

「内容の伝達レベルの評価」に関しては、3 名の評価者による重複評価の結果、3 評価者間で十分に整合した相対順位が得られた。このことから、「内容の伝達レベルの評価」は、現基準にて十分に相対評価に実用できることが検証された。

しかしながら、絶対評価の精度に大きく影響する評価水準（評価の厳しさ、甘さ）に関しては、3 評価者間で大きなずれが生じており、評価水準を揃えるための方策が必要であることが判明した。このような評価水準のずれが生じた最大の原因として、今回の品質評価実践では、評価者間の意識合わせが不十分であったことが挙げられる。これを踏まえ、次章にて具体的な意識合わせの手法を検討し、絶対評価を行う場合には、この手法を要件化することとする。

「重要技術用語の翻訳精度の評価」「流暢さの評価」「チェックリストに基づく評価」に関しては、今回の品質評価実践により、それぞれ破綻なく評価作業が実施できることが確認できた。このうち「チェックリストに基づく評価」については、フィードバック利用が主目的であり、単独のシステムが対象となる場合が多いと想定されるが、システムの翻訳ミスの傾向や弱点を把握するためには、他のシステムを同時に評価して結果を比較することが有効である。

8.7.3. 相対評価方法

今回の品質評価実践の結果、上記のとおり、「特許庁手法」に定められた翻訳観点・基準及び評価対象文数は、複数の翻訳エンジンの優劣を比較する相対評価に十分に実用可能であることが検証された。

8.7.4. 絶対評価方法

絶対評価方法に関しては、品質評価実践結果に基づき、現行のマッチング基準の妥当性を検証し、その結果を踏まえて新しいマッチング基準を設定した。ただし、今回の検証は、単一のデータ（1名の元特許庁審査官による判定）に基づいており、また、これと対応させる「内容の伝達スコア」も、その評価水準にずれが検出されていることから、新しい基準の妥当性については、十分に検証されていない。今後、より多数かつ正確なデータを取得して、その妥当性を検証し、基準を調整していく必要がある。

第 9 章

特許庁機械翻訳品質評価手法の改良と今後の課題

第9章 特許庁機械翻訳品質評価手法の改良と今後の課題

「特許庁手法」の実践（第7章）及びそこで得られた評価結果の分析（第8章）、さらにその分析結果を踏まえた第2回委員会¹での妥当性検証により、「特許庁手法」は、少なくとも相対評価には十分に利用できることが確認できた。同時に、これらを通じて「特許庁手法」の課題や懸念、問題点、今後の留意事項等も判明した。以下に、これらについてまとめた。

9.1. 評価精度の向上

「特許庁手法」の評価精度をさらに向上させるための方策について本項にまとめる。

9.1.1. 評価者間の意識合わせ方法の改善

「特許庁手法」を用いた品質評価実践により、複数のエンジン間の相対評価に関しては、同手法が実用に足りるものであることが確認された。しかしながら、その一方で、同手法のメインの評価観点である「内容の伝達レベルの評価」では、3名の評価者間で評価水準に大きなずれが生じていたことが判明した（8.2.3.）。評価水準のずれは、相対評価においては致命的な問題とはならないが、絶対評価が目的である場合には、あらかじめ各評価者の評価水準を十分に整合させることが必須であり、評価者間の意識合わせの方法について再検討し、具体的な方法を定めておく必要がある。

今回の品質評価実践では、評価水準を合わせるための方策として、各観点の評価基準を本文にて明確に定義し、これを詳しく解説した評価手順書や、多様なケースをカバーした評価サンプルを提供することによって評価者が判断に迷わないよう努めたが、その一方で、各評価者間の意識合わせは、それぞれの1日分の評価結果（対象文は不定）を回収し、リファレンスとしてあらかじめ用意しておいた評価結果と比較して、評価値が2以上違うものに対してのみ再確認を促す（修正するかどうかは任意）という、簡易な方式にて行なった。

今回の品質評価実践の結果から、上記方式の意識合わせでは評価者間の評価水準を十分に揃えることができないことが分かったため、より適切な手法を検討し、「特許庁手法」に具体的に定義する。その具体的な手法として、NTCIR-9にて採用された手法を参考とした。

NTCIR-9にて採用された意識合わせの手法²は、「3人の評価者が同一の訓練用の50文（または100文）について評価を実施し、基準を合わせるために評価者が話し合っ

¹ 第2回委員会の主な提言・指摘事項は、添付資料1-5を参照。

² 第2章2.1.2.4.参照。

価結果をこの 50 文（または 100 文）に対して決め、この統一の評価結果を評価の目安として用いる」というものである。なお、NTCIR-9 は各評価者が異なる文の評価を担当する形であったため、このように訓練用の文を別途用意する必要があったが、「特許庁手法」では、相対評価を目的とする場合であれば、評価対象文 200 文中の 50 文を用いて同様の意識合わせを行うこととすれば、評価作業の負担をさほど増加させずに効率的な意識合わせができると考える。

ただし、絶対評価を行う場合は、評価者間で合わせた評価水準が「特許庁手法」の意図した通りの水準となるよう、より入念に意識合わせを行う必要がある。具体的には、まずは各評価者が、正解データが存在する訓練用の文を（正解データを見せずに）30 文ほど評価し、その後、提示された正解データに基づいて自らの評価水準を意図されたレベルに修正したうえで、あらためて実際の評価対象文からさらに 30 文程度を評価し、その結果を持ち寄って評価者間で話し合い、統一の評価結果を決めて、評価作業を管理する立場の者の承認を得る、という手順を踏むべきである。なお、評価作業を管理する立場の者は、各評価者の評価結果と統一の評価結果とのずれを確認し、必要に応じて、各評価者を個別に指導して評価水準を調整するとよい。ここで得た統一評価結果は次回以降の正解データにも利用できる。

現行の「特許庁手法」は、「内容の伝達レベルの評価」部分において、意識合わせについて「＜評価に先立って、サンプルデータ及びサンプル評価の意識合わせを行う。なおサンプル評価結果は、意識合わせの内容を反映させたいと、次回以降のリファレンスとして保存・活用する＞」との注記を記載していたが、上記を踏まえ、これを以下のとおり具体的に定義し、「絶対評価」を目的とした評価を行う場合の必須要件とした。

＜評価に先立って、サンプルデータ及びサンプル評価の意識合わせを行う。特に、「絶対評価」を目的とする場合には、「相対評価」の場合と比べて入念な意識合わせを行う。具体的には、各評価者が、正解データが存在する訓練用評価文 30 文程度にて評価を試行し、評価結果と正解データを照合して自らの評価水準を意図されたレベルに修正したうえで、実際の評価対象データから選定した 30 文程度について、各評価者が評価結果を持ち寄り、統一の評価結果を話し合いにより定め、評価作業管理者の承認を得た後、これを目安として残りの評価対象文の評価を行う。なおサンプル評価結果は、次回以降のリファレンスとして保存・活用する＞

「内容の伝達レベルの評価」のような主観的評価では、このような入念な意識合わせを行なったとしても、複数の評価者の評価結果を完璧に揃えることは非常に難しい。しかしながら、こうした意識合わせを実施したうえで複数名による重複評価を行うことにより、各評価者の評価結果を平均した評価スコアにおいては、十分に安定した評価水準を得ることが期待できる。

9.1.2. 正解データの充実

今回、「内容の伝達レベルの評価」において評価水準にずれが生じた一因として、リファレンスとして提供した評価サンプルが不十分であった可能性があり、サンプルを更に拡充する必要がある。

評価サンプルは、いわば評価の正解データといえるものである。評価の正解データは、多数作成しておけば、評価サンプルの拡充以外にも、例えば 9.1.1.にて行う事前意識合わせのリファレンスや、9.3.にて後述する絶対評価スコアとのマッチング等にも利用できる有用なデータとなる。本調査では、この考えに基づき、今回の品質評価実践結果データより「正解データ」を抽出することとした。

「正解データ」の抽出条件は、リファレンス評価（8.2.2.参照）における「内容の伝達レベル」の評点を基準とし、3 名中 2 名以上の評価者の評点がこれと合致し、かつ、その評点及び「重要技術用語」の適訳率から導出される絶対評価スコア（新マッチング基準による）が、元審査官の付与した検証スコア（8.6.2.参照）と合致するものとした。その結果、995 文中 163 文³が「正解データ⁴」として抽出された。表 9-1 に、その構成を示す。

表 9-1. 「正解データ」の構成

	スコア 1	スコア 2	スコア 3	スコア 4	スコア 5	合 計
RBMT1	9	14	3	4	1	31
RBMT2	7	9	5	10	3	33
RBMT3	3	3	7	9	2	24
SMT4	2	11	2	5	20	40
SMT5	7	6	8	4	9	33
合 計	28	43	25	32	35	163

これら「正解データ」の原文は、全て英日テストセットに含まれるものである。このため、今後「正解データ」と同じ原文が評価対象文となる可能性があるが、そのこと自体に大きな不都合はないと考える。

なお、この「正解データ」をサンプルとして評価者に提供する際には、各スコアあたり 2 ～3 文程度を厳選して提供する⁵。あまりに大量の評価サンプルは、評価者がそれを確認する作業に多大な労力を要し、評価作業全体としての効率を悪化させる要因となりかねない。

³ 原文は 117 文。1 文に対し複数の機械翻訳結果が採用されている場合がある。

⁴ 詳細は「正解データ.xlsx」（電子データ）を参照。

⁵ 添付資料 7-3 の P.6【評点 5 のサンプル集】以降が、今回「正解データ」から追加した評価サンプルとなる。

また、今回の品質評価結果で「正解データ」とならなかったもの、すなわち、リファレンス評価と各評価者の評価値が整合しなかったものも、評価者の意識合わせ用の訓練データとして活用することが可能である。特に、リファレンス評価と評価者の評価値が大きく食い違っているデータは、評価のゆれが生じやすいケースであるため、これらをリファレンス評価の評価根拠を説明する形でサンプルとして提示することは有効である。

9.1.3. 評価基準への補足

「特許庁手法」を用いた品質評価を実践した結果、各評価観点において、現行の評価基準や評価手順書、評価サンプルでは、どのように評価すべきか明確に定義されていないケースがいくつか見られた。このことも、「内容の伝達レベルの評価」における評価者間のずれの一因となった可能性がある。

本節では、各観点において評価者が判断に迷った実例を示し、それぞれについて「特許庁手法」で以後どのように判断すべきかを定めた。なお、ここで定めた内容は、評価手順書、評価サンプルに反映させた。

9.1.3.1. 内容の伝達レベルの評価

はじめに、「内容の伝達レベルの評価」に関する補足事項について示す。

〔文の骨格構造に関する誤り〕

「内容の伝達レベル」の評価対象は、原文に含まれる全ての技術的要素とその相互関係であり、これらが機械翻訳結果に正しく伝達されているかを、それぞれの重要度と訳語の精度から、総合的に評点を判断することとしている（4.3.1.4.参照）。

したがって、機械翻訳結果において文の骨格構造が大きく誤っている場合でも、無条件で最低評価（評点 1）とはならず、個々の構成要素等に正確に訳されているものがあれば、それらに応じて評点 2 以上が付与される場合もある。ただし、この「文の骨格構造」の重要度は明確に定義していなかったため、品質評価実践後、評価者より、文の骨格構造が大きく誤っていた場合の判断が困難であったとの指摘を受けた。

この指摘を踏まえ、文の骨格構造の重要度について、以下のとおり整理し、これを「内容の伝達レベルの評価」の評価サンプル（添付資料 7-3）に反映させることとした。

文の骨格構造が大きく誤っており、文脈や構成要素等を踏まえても本来の文の主旨を類推不可能である場合は、評点は 2 以下とする。また、読み手に文の主旨を誤解させる可能性の高い誤りは、程度によっては「(本来の) 文意がわからない」ものとし、構成要素が適切であっても、評点 1 としてよい。

具体例は以下のとおり。

原文	Provided is an emission treatment system and method for simultaneously remediating the nitrogen oxides (NO _x), particulate matter, and gaseous hydrocarbons present in diesel engine exhaust streams.
人手翻訳	ディーゼルエンジン排気流れに存在する窒素酸化物（NO _x ）と粒状物と気体状炭化水素を同時に修正するに適した 排気処理システムおよび方法を提供する。
機械翻訳	排気処理システムおよび方法は、同時に窒素酸化物（NO _x ）を是正するため、粒子状物質、及びディーゼルエンジン排気流中に存在する ガス状炭化水素を提供する。

上記例は、技術用語は概ね適切に訳されているものの、文の骨格構造は完全に誤っており、発明の主題が「排気処理システムおよび方法」ではなく「ガス状炭化水素」とであると誤解される訳文となっている。このような文は、「骨格構造が大きく誤って」おり、「本来の文の主旨を類推」することが困難、かつ「読み手に文の主旨を誤解させる可能性」が非常に高いため、評点は1とする。

このような文も、技術用語が正しく訳されていれば検索用途には有用であることは事実である。しかし、技術用語の翻訳精度は「重要技術用語の翻訳精度の評価」でも評価され、絶対評価にも、この観点の評価が一定値以上であればスコア 2（粗読又は検索データに適した翻訳品質）が保証されるよう改めている。このため、ここで技術用語の訳質を過度に反映させる必要はないと判断した。

〔誤解を招く表現〕

語単位であっても、明らかな誤訳でなく、読み手がそれと気づきにくいような誤訳は、読解時の悪影響が大きいと、「訳語の精度」が通常の誤訳よりも悪いものとして、より大きく減点することとした。

原文	Non-ferromagnetic layer 104 may include, for example a conductive material layer having a thickness in a range of 1-3 nm, although non-ferromagnetic layer 104 may be thicker or thinner, in other embodiments.
人手翻訳	非強磁性層 104 は、たとえば厚さが 1～3 nm の範囲の導電性材料層を含み得るが、他の実施例では非強磁性層 104 はより厚い場合もあれば、より薄い場合もある。
機械翻訳	非強磁性層 104 は、他の実施形態では、より厚い又はより薄いかもしれないが、非強磁性層 104 は、例えば、 1～20nm の範囲の厚さを有する導電性材料層を含むことができる。

上の例は、数値範囲「1～3nm」が「1～20nm」と誤訳された事例である。原文と照らさない限り、実用時に読み手がこのような誤訳に気づく可能性は低く、大きな誤解を与える

危険性が高い。この機械翻訳結果は、文の骨格構造も正しく、各構成要素の訳語も適切であり、誤訳はこの数値範囲部分だけであるが、この誤訳は重大であり、少なくとも文意を20%以上損ねていると判断して、評点は3とする。

[特許知識により正確に補正できる誤訳]

評価者が自身の特許知識・経験により正確に補正できる誤訳については、補正した内容にて評価することとした。

原文	13. The method of claim 1 , wherein the powder coating composition comprises a polymer selected from the group consisting of an acrylic, polyester, polyurethane, polyether, polyvinyl, cellulosic, acrylate, silicon-based polymers, co-polymers thereof, and combinations thereof.
人手翻訳	【請求項13】 前記粉末被覆組成物が、アクリル、ポリエステル、ポリウレタン、ポリエーテル、ポリビニル、セルロース、アクリレート、ケイ素系ポリマー、これらのコポリマー及びこれらの組合せからなる群から選択されるポリマーを含有する、 請求項1 記載の方法。
機械翻訳	13. 請求項1 (粉末コーティング組成物がアクリル、ポリエステル、ポリウレタン、ポリエーテル、ポリビニル、セルロース誘導体、アクリレート系繊維、シリコン系ポリマー、そのコポリマーおよびそれらの組み合わせから成る群から選択されるポリマー) の方法。

上記例の原文は、請求項1を引用した引用請求項だが、機械翻訳結果は、引用請求項の内容を請求項1の内容として扱っており、大きな誤解を生む。ただし特許文献に通じている者にとっては、引用請求項にて被引用請求項の内容のみが示されることはありえないことは自明であり、これを正しく補正して解釈することが可能である。このため、機械翻訳結果の文字通りの内容は正確だが、それを補正して得られる情報は正確である。

こうした補正を許容すると、個々の評価者の特許知識により評価基準にずれが生じる要因となることは事実である。しかしながら、評価の目的を考えると、機械翻訳結果の正確性よりも実用性が重視されるべきであり、実用時に可能な補正は有用性評価にも反映させるべきと判断した。

9.1.3.2. 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の評価」の実践を通じて判明した課題を踏まえた補足事項について示す。

〔原語のカタカナ表記〕

今回の品質評価実践を通じ、原語をカタカナ表記しただけの訳語をどのように評価すべきかについて、明確な基準が必要であることが判明した。

例えば「melting：溶解」が「メルティング」と訳されているような場合、日本語で一般に使われる同義語の範囲ではないため「適訳語（A）」とはなりえないが、誤った意味に訳されているわけではないため「誤訳語（C）」とすることはできず、現状の評価基準では、常に「可訳語（B）」とせざるをえない。類似の例として、例えば以下のような実例が存在した。

gas transfer	ガス・トランスファー
stochastic screen pattern	ストカスティックなスクリーンパターン
SHEET FOLDING METHOD	シートフォールディング法
electric field	エレクトリックフィールド

原語が英語であれば、上記のようなカタカナ語の多くは、専門家であれば意味は正しく把握できると思われる。このため、現行の基準どおり「可訳語（B）」とするのが妥当である。だが、仮にこれらが英語以外の語のカタカナ表記で出力されていたとしたら、専門家であっても意味の把握は難しい場合も想定される。今後、こうしたものも「可訳語（B）」として取り扱うべきかどうかについて改めて検討した。

本評価の評価者は当該技術に通じた専門家であり、彼らが読んで意味が分からないカタカナ語は、厳密な意味では誤訳とは言えないにせよ、実質的な有用性は「誤訳語（C）」や「不訳語（D）」と大差がない。したがって、結論としては、評価者が読んで意味の分からない訳語は、カタカナ表記であっても誤訳（C）とみなすこととした。

9.1.4. 流暢さの評価

[内容のナンセンスさの反映]

流暢さの評価において、内容のナンセンスさを減点対象とすべきか、もしくは、日本語として純文法的に自然であれば、技術内容的に異常であっても減点対象とすべきでないかを、明確に定義することとした。

原文	[0009] FIG. 3 is a top view of a lid of the microwave steamer of FIG. 1
人手翻訳	[0009] 図3は、図1のマイクロ波蒸し器の 蓋部 を示す上面図である。
機械翻訳	[0009] 図3は、図1のマイクロ波汽船の 眼瞼 の平面視である。

上例の機械翻訳結果は、文の構造として文法的には全く問題がない。また、「steamer : 蒸し器」が「汽船」と誤訳されていることも、「流暢さの評価」では機械翻訳結果を独立した文として見るため、問題とはならない。だが、本例ではさらに「lid : 蓋部」が「眼瞼」と誤訳されており、こちらは、「汽船」とは技術的に相容れない用語であるため、機械翻訳結果のみを見ても、この文の内容が正しくないことは明らかである。このような「文法的には問題がないが技術内容的にナンセンス」な文を、「文の理解し易さ」の観点から、減点対象とすべきかどうか問われた。

結論としては、「特許庁手法」では、機械翻訳結果にどのような語が出現したとしても、それが文として成立しており、文法規則に沿っている限り、「流暢さの評価」では減点対象とはしないこととする。

「流暢さの評価」では、文法的に正しくても技術内容的にナンセンスな文は低い評価にすることが多いことは事実だが、「特許庁手法」は、複数の評価観点を組み合わせて総合評価を行う想定であり、上例のような文は、「内容の伝達レベルの評価」等で低い評価が下される。このため、ここでは文としての「文法的な」自然さのみを評価対象とするのが妥当であると判断した。

9.1.5. チェックリストに基づく評価

〔自動詞・他動詞の誤り〕

品質評価実践時の「特許庁手法」の評価基準および評価手順書では、動詞の自動詞・他動詞の誤りを、「チェックリストに基づく評価」の【必須項目】の「D. 品詞誤り」と「F. 語の誤訳」のどちらにカウントすべきかが明確でなかった。

原文	The interactive element 220 is illustratively configured to be positioned in close proximity with the data bits 210 and the second data track 204 to read and/or write the information on each of the individual data bits 210 as the data medium 200 and the interactive element 220 move relative to each other.
人手翻訳	例示のとおり、インタラクティブ・エレメント２２０はデータビット２１０および第２のデータトラック２０４にごく接近して位置決めされ、データ媒体２００およびインタラクティブ・エレメント２２０が互いに 移動する と個別データビット２１０の各々で情報を読み書きするように構成される。
機械翻訳	双方向素子２２０は、データビット２１０及び第２のデータトラック２０４と、近接読み出し及び／又はデータ媒体２００と双方向素子２２０を相対的に 移動する ときに、個々のデータビット２１０の各々に情報を書き込みに位置決めするように模式的に構成されている。

move は上例では自動詞（data medium 200 と interactive element 220 が自ら動く）として使われているが、機械翻訳では他動詞として訳されている。このような場合、チェックリストの【必須項目】において、これを「D. 品詞誤り」とみなすか、もしくは「F. 語の誤訳」と見なすか、現行の基準では評価者によって判断が分かれる可能性がある。

これを避けるため、今後「特許庁手法」では、自動詞・他動詞の誤りは、一律「F. 語の誤訳」として扱うこととした。自動詞・他動詞の誤りは、少なくとも動詞としては扱われているため、文構造への影響も（他の品詞誤りに比べると）軽微であることが多いため、そのように判断した。

9.2. 相対評価方法における課題

本項では、「特許庁手法」の品質評価結果を用いた相対評価における課題をまとめた。

9.2.1. システムの定点観測

「特許庁手法」では、品質評価の目的・利用方法のひとつである「相対評価」の対象として、複数システム間の翻訳精度比較だけでなく、「特定のシステムにおける翻訳品質の定点観測（同一システムの二つの時点の翻訳精度比較）」への利用も想定している（4.1.3.）。本調査にて実践した品質評価は複数システム間の翻訳精度比較であり、特定システムの定点観測に関しては未実施であるため、ここでその手法について検討する。

特定システムの定点観測であっても、新旧の機械翻訳結果をその都度作成し相対評価する場合は、複数システム間の翻訳精度比較と大きな違いはない。ただし留意すべき点として、同一のシステムであるため、機械翻訳結果も同一のものが出力されるケースが多いという点が挙げられる。相対評価においては、まずは双方の機械翻訳結果を比較し、双方に違いがあるもののみを人手評価するのが効率的である。また、システムによっては、新旧システムが併存しない（つまり、新システムがリリースされた時点で旧システムの利用が不可能となる）場合もありえる。このような場合、あらかじめ旧システムの機械翻訳結果を保存しておく必要がある。具体的には、テストセット全文の機械翻訳結果をあらかじめ作成しておくべきであろう。

一方、システムの定点観測として考えられるもう一種の方式として、旧システムについては既存の評価結果を使用し、新システムのみを新たに評価して、その結果を比較する方式が挙げられる。この方式を採る場合、さらにいくつか留意すべき事項がある。

まず、評価対象文については、新旧システムで同一の文を用いるべきである。異なる評価対象文で難易度を完璧に合わせることは不可能であり、難易度の異なる評価対象文にて行なった評価結果では、その差異がシステムの性能差によるものか、対象文の難易度の差によるものかが判断できない。特に同一システムの定点観測の場合、双方のシステムの性能差はかなり小さい場合も考えられ、評価対象文の難易度の違いに吸収されてしまう懸念がある。

また、新旧システムを同時に評価しない場合、同一の評価者を起用できるとは限らず、仮に同一の評価者を起用できたとしても、時間が経過しているため前回と同一の評価水準が維持されない懸念がある。したがって、評価者間の評価水準のずれに対する一層の配慮が必要である。本来、「特許庁手法」では、1文に対し複数の機械翻訳結果を評価する場合、同一の評価者が同時に評価することを要件としており、定点観測の都度、新旧システム双方の機械翻訳結果を同一の評価者を起用して相対評価することを推奨するが、これが不可能であり、新システムの出力結果のみを評価せざるを得ない場合は、評価者に旧システムの機械翻訳結果とその評点を提示することが有効である。評価者は、旧システムの機械翻訳結果と、それに対して付与された評点を参照し、これとの比較で新システムの評価を行

うことができるため、正確な相対評価が可能となり、また評価水準もおのずと前回の水準と揃うこととなる（前回の評点を提示しておけば、前回評価時よりも改善している出力に前回よりも低い評点は付与されない）。

最後に、同一システムの定点観測を行う場合に特に留意すべき点として、新旧システムで出力に変化が生じている場合でも、必ずしも評点を変化させる必要はないという点を、評価者に徹底しておく必要があることが挙げられる。新旧システムの相対評価では、両者の出力に変化が生じている場合、わずかな改善であっても評点に差をつけるべきという考え方もある。だが、定点観測が定期的に行われることを想定すると、わずかな改善に対して前回よりも高い評点を付与すると、次回以降に再度わずかな改善があった場合、さらに高い評点を付与することとなるため、実際の翻訳品質よりも過度に高い評点となってしまうという事態が生じかねない。これは、評価結果を絶対評価にも利用する場合には特に問題となる。

9.2.2. 相対順位の有意性が充分でない場合の対処策

「特許庁手法」の定める 200 文という評価対象文数は、「内容の伝達レベル」及び「重要技術用語の翻訳精度」の観点におけるエンジン間の相対比較にはほぼ十分な件数であることが、それぞれの評価結果に対する符号検定（8.2.4.、8.3.3.）により検証された。ただし、その一方で、性能が極めて接近しているエンジン間の比較では、200 文の評価では有意性のある相対順位を得ることができない場合もあることもわかった。

このような場合に両者の優劣を定める方法として、評価対象文数を増やしたうえで、性能の接近したエンジンの翻訳結果のみに対して、文単位にどちらが優れているかを相対評価することが推奨される。5 段階評価では同じ評点となるような小さな訳質の差も反映した比較結果を得ることができ、かつ評価者にとって 5 段階の評価値を選択する評価よりも作業負担が小さい。

ただし、機械翻訳結果の品質は、仮にひとつの評価観点において優劣がつかなくても、他の評価観点を含めた総合評価にて定められることができれば充分である場合が多い。さらに、エンジンの優劣の判断は、「特許庁手法」で測れる翻訳品質のほかにも、価格やインタフェース、チューニング・カスタマイズの容易さ等、多角的に行うべきであり、品質評価の一観点における優劣を常につける必要があるかを、評価目的等に鑑みて十分に検討すべきであろう。

9.2.3.「流暢さの評価」におけるサンプル 50 文評価の妥当性

「特許庁手法」では、「流暢さの評価」の対象文数は、作業負担軽減の観点から、「内容の伝達レベル」等の評価対象文 200 文からサンプリングした 50 文のみに対して行うよう定めた。しかしながら、8.4.4.に示した符号検定の結果、サンプリング 50 文にて得られる相対順位は有意性が不十分な場合が多く、他の評価観点と同等の有意性を得るためには、やはり 200 文全文に対して評価を行う必要があることが判明した。

「流暢さの評価」は、評価観点としては補助的なものであり、「内容の伝達レベルの評価」や「重要技術用語の翻訳精度の評価」と比べると、その重要性は低い。サンプリング 50 文としたのはそのためであり、評価作業の負担軽減を優先した結果である。しかしながら、今回の品質評価実践により、「流暢さの評価」において最も作業負担が大きいのは機械翻訳結果の読解作業であり、この作業自体は、その後に行う「内容伝達レベルの評価」でも必要な作業であるため、「流暢さの評価」を省略してもさほど大きな負担軽減とはならないということがわかった。

これを踏まえ、「特許庁手法」では、以後「流暢さの評価」についても 200 文全文を評価対象とすることに改め、以下の定義を廃止した。

~~＜機械翻訳対象文のうちランダムに選択した 50 文のみを対象とするサンプリング評価を行う。一定数の長文を含める＞~~

9.3. 絶対評価方法における課題

「特許庁手法」では、機械翻訳結果の各種特許業務への有用性を品質評価結果から直接判断するための方法及び基準を「絶対評価方法」として定義している。今回の品質評価実践では、その定義内容の妥当性を検証し、新たなマッチング基準を定めた（8.6.参照）。

表 9-2 に、上記過程を経て「特許庁手法」にて定義された絶対評価方法の内容を示す。

表 9-2. 「特許庁手法」の「絶対評価方法」部分

3. 絶対評価方法

(0) 絶対評価スコアと用途とのマッチングは、以下のとおりとする。

スコア 5：権利認定、引用発明認定に適した翻訳品質

スコア 4：実体審査、先行技術調査における精読に適した翻訳品質

スコア 3：先行技術調査及び文献照会時の粗読及び検索データに適した翻訳品質

スコア 2：粗読又は検索データに適した翻訳品質

スコア 1：特許調査に際しては全く利用できない翻訳品質

(1) 絶対評価スコアを、「内容の伝達レベルの評価」のスコア及び「重要技術用語の翻訳精度の評価」の適訳率に基づき定める。

内容伝達 適訳率	1.0～3.0	3.0～3.5	3.5～4.5	4.5～5.0
60%以上	2	3	4	5
60%未満	1	2		

(2) 「流暢さの評価」において、特筆すべき問題点が存在する場合は、その内容に応じてスコアを減点する。

しかしながら、この新たなマッチング基準は、第 8 章 8.7.4.にて述べたとおり、単一のデータ（1 名の元特許庁審査官による判定）に基づいて定めたものであり、これと対応させる「内容の伝達スコア」も、その評価水準にずれが検出されており、精度が充分でない可能性があることから、その妥当性は十分に検証されていない。したがって、今後さらに検討・検証を行い、その内容を整備していく必要がある。

本項では、「特許庁手法」に定めた現行の「絶対評価方法」における課題及び問題点についてまとめる。

9.3.1. 「内容伝達レベルの評価」の精度の問題

9.3.に示した現行の絶対評価方法では、各種特許業務に対応する「絶対評価スコア」を決定するために、これと「内容の伝達レベルの評価」の評価スコア及び「重要技術用語の翻訳精度の評価」の適訳率とを照合させるためのマッチング基準を(1)のとおりに定めている。このマッチング基準は、本調査にて実施した品質評価の結果に基づいて調整したものである。だが、今回の品質評価実践では、9.1.1.でも述べたとおり、「内容の伝達レベルの評価」において、重複評価を行なった 3 名の評価者間で評価水準のずれが生じており、この評価スコアに基づいて定めたマッチング基準の信頼性について疑いが持たれた。

しかしながら、今回の品質評価実践に用いた評価スコアは、3 評価者間でばらつきはあったものの、これを平均した結果は、事前に想定していた水準（リファレンス評価の平均スコア）から大きな逸脱はなかった（8.2.2 参照）。したがって、これを根拠に定めた新マッチング基準も、今後の評価実践において、評価水準が常にリファレンス評価の水準から大きく逸脱しないよう適切な方策を採れば、一定の妥当性は有すると考える。

とはいえ、今回の品質評価実践で 3 評価者の評価水準に大きなずれが生じていたことは事実であり、平均スコアが事前に想定した水準に近似したことは偶然に近い。このため「特許庁手法」では、9.1.に示したような評価精度の向上策を講じ、今後の運用において内容伝達スコアの評価水準が今回新マッチング基準の根拠としたものと常に近似するよう配慮する。

9.3.2. 検証用データの精度の問題

本調査では、絶対評価方法のマッチング基準の妥当性検証のため、元特許庁審査官を評価者に起用し、品質評価実践の対象とした全ての機械翻訳結果（199 文×5 エンジン）について、各文の絶対評価スコア（＝検証用スコア）を直接取得した（8.6.1.～2.）。「特許庁手法」の現行のマッチング基準は、この検証用スコアに基づいて定められたものである。

しかしながら、この検証用スコアはあくまで 1 名の元審査官の判定に基づくものであって、これが審査官の平均的な判定であるとはいえない。本質的に、どの程度の品質の機械翻訳結果を特定の用途に「有用」と判断するかには、少なからぬ個人差があると考えられ、本来こうした検証は、少なくとも複数名のサンプルを取得したうえで行う必要がある。

また、本調査にて実施した検証方法は、文単位で行われており、評価者は「文献全体がこの文と同等の精度で翻訳されていると仮定した場合の有用性」を判定している。個人の想像力に委ねる部分の大きい方式であるため、個人差はさらに大きくなる。

なお、特許文献の機械翻訳結果は、通常は文献単位で利用するものであり、本来、有用性の判定は、文献単位で行うことが望ましい。しかしながら、その場合、1 つの評点を得るために 1 文献全体を読解する必要が生じ、過大な作業負担が生じるため、現実的な手法と

はいえない。

したがって、検証スコア取得の方法としては、本調査の実施した方法が妥当であると考ええる。ただし、今後こうした検証を行う場合は、評価の主旨を十分に説明したうえで、具体的な評価基準を示すサンプルを提示し、さらに「内容の伝達レベルの評価」と同様のサンプル評価による意識合わせを実施することが望まれる。

9.3.3. 絶対評価スコアの定義に関する課題

「特許庁手法」では、絶対評価スコアの定義自体は、「仮設手法」のものをそのまま流用している。その内容は以下のとおりである。

スコア 5：権利認定、引用発明認定に適した翻訳品質
スコア 4：実体審査、先行技術調査における精読に適した翻訳品質
スコア 3：先行技術調査及び文献照会時の粗読及び検索データに適した翻訳品質
スコア 2：粗読又は検索データに適した翻訳品質
スコア 1：特許調査に際しては全く利用できない翻訳品質

しかしながら、このうちスコア 5 とスコア 4 は、元審査官経験者からも、それぞれの用途において有用となる機械翻訳結果にどのような品質差があるのかを想像できない、との指摘があった。絶対評価スコア自体は、評価者が直接各文に付与するものではないとはいえ、各スコア、特にスコア 4 と 5 に該当する翻訳品質の違いがどのような点にあるのかを明確にしておかないと、今後の検証も精度を欠くこととなる。

本調査にて実施した検証作業においては、この部分は評価者（元特許庁審査官）の主観に委ねており、スコア 5 と 4 とを明確に区別できるような基準（例えば、係り受けに 1 つでも誤りがあればスコア 5 はつけない等）は示していない。後の取材では、機械翻訳結果が訳文として完璧であればスコア 5、若干でも不備があればスコア 4 以下としたとのことであったが、この判定基準が審査官の平均的な考え方であるかは未確認である。

したがって、スコア 5 とスコア 4 との品質差を明確にするためには、多数の審査官による検証作業を行い、判定結果を分析しつつ、どのような判断基準により 4 と 5 を区別したかを各審査官に取材する必要がある。

ただし、こうした作業は大きなリソースを必要とするものである。多大なリソースを費消してスコア 4 と 5 を区別する必要があるかをまず検討する必要があるであろう。現状の機械翻訳技術では、200 文程度を対象とした内容伝達スコアで、絶対評価スコア 4 以上に到達するレベルの機械翻訳結果が得られる可能性は低い。今回の品質評価実践でも、多くのエンジンがチューニング前の状態であったとはいえ、内容伝達スコアは最高で 3.15、平均

2.82 と、スコア 4 の最低ラインである 3.5 を大きく下回っており、特段に機械翻訳と相性のいい言語対でない限り、当面、スコア 4 に到達するようなケースは期待しにくい。

このことから、絶対評価スコアの妥当性検証も、まずはスコア 1～3 の条件設定に注力するのが効率的であろう。スコア 1～3 であれば、それぞれの基準設定の考え方は 8.6.3. に示したマトリックスが有効であり、これをベースとして、閾値の設定を、より信頼性の高いデータで見直していくべきである。

9.3.4. 今後の方向性

本調査にて行なった絶対評価スコアのマッチング基準の妥当性検証 (8.6.) は、検証に用いた品質評価結果及び検証用スコアの精度が十分に保証されていないため、これに基づいて定めた新マッチング基準自体の信頼性が低いことは事実である。しかしながら、この妥当性検証の方法自体はおおむね正当なものであり、精度の検証された数値を用いることで、信頼性の高い新マッチング基準が設定可能であると考ええる。

ただし、今回の検証方法では、「粗読又は検索データに適した翻訳品質」を示す「絶対評価スコア 2」をそのまま判定に利用した結果、スコア 2 に該当する翻訳結果が「粗読」と「検索」のどちらへの有用性によるものかを後で判別することができず、この部分において十分に精緻な分析ができなかった。したがって、今後の留意点として、検証の際には、スコア 2 を「粗読」と「検索」とに分けて判定することが挙げられる。

また、絶対評価スコアの定義については、現状の機械翻訳のレベルでスコア 4 以上の品質が得られることは稀であり、スコア 1～3 について重点的に検証を行うのが効率的である。

9.4. テストセットに関する課題

「特許庁手法」に定めたテストセットの定義に関する課題や留意点について本項にまとめる。

9.4.1. パテントファミリー対訳文の利用

テストセットの作成方法について、「特許庁手法」では、理想的にはパテントファミリー（対応特許）不在の特許文と新規に作成した人手翻訳を用いるべきとしつつも、コスト面を鑑みて、パテントファミリー対訳文の利用も許容している（4.4.1.2.）。本調査で作成した英日テストセットも、全てパテントファミリー文献から対訳文を抽出して作成している。

しかしながら、パテントファミリー対訳文は一般に入手可能な既存データであるため、例えば SMT 方式の機械翻訳エンジンにおける学習データ等、エンジンのチューニングに利用される可能性がある。この場合、テストセットに含まれるパテントファミリー対訳文をチューニングに利用したエンジンが、不当に有利な条件で品質評価されることとなり、評価の信頼性が低下する。事実、今回の品質評価実践でも、評価対象とした一部エンジンにおいて、評価対象文がチューニングに利用された形跡が見られた（9.4.3.にて後述）。

SMT 方式の機械翻訳エンジンは既に各所で実用化されており、「特許庁手法」を用いた品質評価においても、今後常に評価対象に含まれることが予想される。このことを考えると、現実的には、テストセットは全文、目的言語で記載された対応文献（英日機械翻訳であれば日本語文献）の存在しない文献から抽出し、参照文を人手翻訳にて作成することが強く推奨される。

現状の SMT 技術では、原語（英日翻訳であれば英語）で記載された原文のみを翻訳品質の向上に利用することは容易ではなく、上記手法であれば、各種方式のエンジンを平等な条件で評価することができる。

なお、SMT 方式といえども商用システムでは学習データの管理が複雑化するため、ここから特定のデータだけを削除することは困難である。そのため、テストセットに含まれる対訳文を学習データから除外するという手法は、実現性に乏しい。

9.4.2. テストセットの再作成

テストセットは、一つのものを恒久的に使用可能な性質のものであるか、それとも定期的に再作成が必要なものであるかについても検討した。

基本的には、文法や特許特有の表現等は頻繁に変化するものではないため、この点に関して、定期的な更新を行う必要性は少ないと考えられる。また、9.4.1.で述べたとおり、テストセットは人手翻訳にて作成することが強く推奨されるが、これに要するコストや、各翻訳言語に対してこれを個別に作成することを考えると、テストセットを定期的に更新す

ることは現実的に難しいと思われる。

ただし、テストセットを恒久的に使用する場合、いくつかの課題への対処が必要となる。まず、技術用語に関しては、テストセット作成以降に一般化した新語の翻訳精度を測ることが不可能となる。だが、RBMT 方式であれば、技術用語の翻訳精度は必ずしもテストセットで評価する必要はない（その語のみを翻訳することでも検証可能）し、SMT 方式の場合は、文中の前後の語に応じて訳語を決定する方式であるため、新語を含む文を用いて新語への対応のみを別途評価するという方法も考えられる。テストセットを用いた評価を前提とする場合も、主要な新語を含む文を追加していくのが妥当であり、テストセット全件を再作成する必要はないと考える。

2 つめの課題としては、テストセットをフィードバック利用目的で使用した場合に、評価に使用した文をエンジン開発者に開示してしまうと、仮にその文の翻訳ミスが解消したとしても、それが根本的な改善によるものか、その文のみにピンポイントで対応した結果なのかを、再度その文を用いて検証することが難しくなる。このため、テストセットはあくまでエンジンの問題点や翻訳ミスの傾向を検出するために用いるものとし、フィードバックには、同様の翻訳ミスが生じる別の文を用いる等の配慮が必要である。

9.4.3. 評価対象文が学習データに含まれている可能性の検証

本調査にて作成した英日テストセットは、全文、英日パテントファミリー文献から採取した対訳文から成る。本調査にて実施した品質評価実践も、この英日テストセットから選定した 199 文を評価対象としているため、全ての評価対象文がパテントファミリーを有する文である。

パテントファミリー文は、一般に入手可能な対訳文であるため、SMT 方式のエンジンの学習データ等、エンジンのチューニングに利用される可能性がある。評価対象文であるパテントファミリー文を、評価対象のエンジンが学習済みであった場合、不当に有利な条件となるため、評価結果の正確性や公平性が失われる。

本調査の品質評価実践でも、一部のエンジンにおいて、評価対象文の少なくとも一部が学習データに含まれていた可能性があることから、以下に示すと通りの検証を行なった。

9.4.3.1. 「重要技術用語の翻訳精度の評価」の結果に基づく検証

まず、今回の品質評価実践で取得した「重要技術用語の翻訳精度の評価」の結果を示す。

表 9-3. 品質評価実践（199 文）による評価結果

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
評価 A (適訳語)	273 語 52.1%	317 語 60.5%	348 語 66.4%	443 語 84.5%	394 語 75.2%
評価 B (可訳語)	108 語 20.6%	87 語 16.6%	90 語 17.2%	41 語 7.8%	46 語 8.8%
評価 C (誤訳語)	136 語 26.0%	110 語 21.0%	83 語 15.8%	39 語 7.4%	61 語 11.6%
評価 D (不訳語)	7 語 1.3%	10 語 1.9%	3 語 0.6%	1 語 0.2%	23 語 4.4%
相対順位 (適訳率)	5	4	3	1	2

(※第 8 章 8.3.2. 表 8-7 を再掲)

次に、「英日機械翻訳の現状分析（第 3 章）」にて実施した、日本語パテントファミリーを持たない 20 文献を対象とした重要技術用語の翻訳精度の評価結果を示す。

表 9-4. 「英日機械翻訳の現状分析」（20 文献）による評価結果

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
評価 A (適訳語)	2,011 語 63.6%	2,117 語 67.0%	2,194 語 69.4%	2,394 語 75.7%	2,264 語 71.6%
評価 B (可訳語)	621 語 19.6%	722 語 22.8%	674 語 21.3%	458 語 14.5%	487 語 15.4%
評価 C (誤訳語)	493 語 15.6%	318 語 10.1%	285 語 9.0%	283 語 9.0%	276 語 8.7%
評価 D (不訳語等)	36 語 1.1%	4 語 0.1%	8 語 0.3%	26 語 0.8%	134 語 4.2%
相対順位 (適訳率)	5	4	3	1	2

(※第 3 章 3.1.7. 表 3-4 に相対順位欄を追加)

これら 2 つの評価結果は、同じ 5 種類のエンジンに対し、同じ評価基準を用いて取得したものである。いくつか条件が異なる部分がある（前者は文単位で 1 語につき 1 回のみ評価、後者は文献全体に対し出現の都度評価、等）とはいえ、非常に類似度の高い評価であると考えられる。事実、5 つのエンジンの相対順位（適訳率）は完全に一致している。

しかし同時に、これら 2 つの評価結果には、RBMT エンジンと SMT エンジンとの顕著な傾向の違いも見られる。具体的には、RBMT の 3 エンジンが、いずれも今回の品質評価実践結果（表 9-3）において、現状分析評価時（表 9-4）よりも適訳率が低下している（-3.0 ～ -11.5 ポイント）のに対し、SMT は、2 エンジンとも適訳率が向上（+3.6 ～ +8.8 ポイント）しており、正反対の傾向が生じている。

こうした傾向の違いは、今回の品質評価実践で評価対象文としたパテントファミリー文が、SMT エンジンの学習データに含まれていた結果生じたものである可能性がある。

今回評価対象とした RBMT エンジンは、いずれもほぼ初期設定で使用しており、パテントファミリー対訳文の有無が技術用語の翻訳精度に影響する可能性は無視できる。したがって、今回の品質評価実践にて各 RBMT エンジンの適訳率が現状分析評価時よりも低下しているのは、今回対象となった重要技術用語の翻訳難易度が、現状分析時のものよりも高かったことを示している。それにもかかわらず、SMT エンジンにおいて今回のほうが適訳率が高いということは、SMT エンジンが前回評価時よりも有利な条件下にあった、すなわち評価対象文の少なくとも一部が学習済みであった、という可能性を示している。特に SMT4 は、前回比 +8.8% と大幅な向上を示しており、評価対象文の少なくとも一部が学習データに含まれていた可能性は高いと思われる。

9.4.3.2. 「内容の伝達レベルの評価」の実施による検証

「英日機械翻訳の現状分析」当時は、「特許庁手法」は作成前であったため、「内容の伝達レベルの評価」は実施されていない。そこで、9.4.3.1. に示した推論、すなわち「SMT4 において、評価対象文の少なくとも一部が学習データに含まれていた可能性が高い」という推論の検証のため、当時の調査に使用したサンプル 20 文献から 2 文ずつ、計 40 文をランダムに抽出し、これらに対して「内容の伝達レベルの評価」を実施した。これら 40 文が SMT の学習データに用いられている可能性はほぼ皆無であり、ここでの評価結果と品質評価実践時の品質評価に顕著な傾向の違いがあれば、その内容によっては、品質評価実践時の評価対象文が学習データに用いられていたという推論を補強するものとなる。

まず、品質評価実践時の「内容の伝達レベルの評価」の評価スコアを表 9-5 に示す。

表 9-5. 品質評価実践時の「内容の伝達レベル」評価スコア

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
スコア	2.39	2.79	3.01	3.15	2.76
相対順位	5	3	2	1	4

続いて、今回実施した 40 文の品質評価結果を表 9-6 に示す。

表 9-6. パテントファミリー不在 40 文の「内容の伝達レベル」評価スコア

	RBMT1	RBMT2	RBMT3	SMT4	SMT5
スコア	2.22	2.47	2.92	2.90	2.66
相対順位	5	4	1	2	3

上表 9-6 のとおり、パテントファミリーを有さない 40 文の評価では、僅差ながら RBMT3 と SMT4 と相対順位が、品質評価実践時のもの（表 9-5）から逆転する結果となった。今回の評価はサンプル数が少なく、また評価者も異なるため、あくまで参考情報の域を出ないが、品質評価実践時の評価結果に比べて SMT4 の相対順位が低下したことは事実であり、先に述べた「SMT4 において、評価対象文の少なくとも一部が学習データに含まれていた可能性が高い」という推論を裏付ける結果と捉えることも可能である。

9.5. 英日以外への言語対への適用可否

本調査では、英日機械翻訳を対象として「特許庁手法」の作成と品質評価実践を行なったが、本手法自体は、言語対を限定せず汎用的に適用できるものとして策定している。この点に関して、第 2 回委員会でも、特段の懸念や課題、問題点は指摘されなかった。

また、本調査では、中日機械翻訳結果（RBMT 及び SMT、計 4 種類）を対象として、「特許庁手法」の各評価観点による品質評価を試行し、問題なく適用可能であることを確認している。

これらより、「特許庁手法」は、英日以外の言語対へも十分に適用可能であると判断できる。ただし、「チェックリストに基づく評価」の【翻訳言語別任意項目】は、本調査では英日機械翻訳に関するもののみを設定しており、他の言語対の品質評価に「チェックリストに基づく評価」を含める場合は、それに先立ち【翻訳言語別任意項目】を設定する必要がある。

9.6. 自動評価

「基本的な考え方」（添付資料 1-1）では、「7. 粗ぶるいについて」の項にて、人手による品質評価に先立ち、自動評価手法を利用して、相対評価に値しないような低品質の機械翻訳結果をあらかじめ検出し、これを除外するという「粗ぶるい」手法について言及している。

本調査では、品質評価実践によって取得した「品質評価結果データ」を用いて、これを主要な自動評価手法のひとつである BLEU、NIST、及び RIBES による自動評価結果と対比することで、上記「粗ぶるい」手法の実用性について実験した⁶。下表 9-7 にその結果を示す。なお、表中「MT0」は、低品質の機械翻訳結果を人為的に作成すべく、各評価対象文を SMT エンジンで英中翻訳し、さらにそれを RBMT エンジンで中日翻訳した結果である。

表 9-7. 「内容の伝達スコアと BLEU、NIST、RIBES スコアの比較

	RBMT1	RBMT2	RBMT3	SMT4	SMT5	MT0
内容伝達スコア	2.39	2.79	3.01	3.15	2.76	—
相対順位	5	3	2	1	4	—
BLEU スコア	0.2640	0.2917	0.3253	0.4543	0.3947	0.1585
	5	4	3	1	2	6
NIST スコア	5.7864	6.1002	6.6208	8.1341	7.4643	4.6428
	5	4	3	1	2	6
RIBES スコア	0.6812	0.7347	0.7492	0.7832	0.7498	0.6025
	5	4	3	1	2	6

上表のとおり、3 種の自動評価全てにおいて、低品質の機械翻訳結果として作成した MT0 が、顕著な差異で最低の評価となっている。この結果からは、自動評価で他よりも顕著に低いスコアのエンジンは、粗ぶるいで除外することは可能であると考えられる。

ただし、その際には、今回用いた 3 種の自動評価手法が、いずれも SMT 方式のエンジンを RBMT 方式よりも高く評価する傾向を示している点への留意が必要である。今回の実験結果では、人手による内容伝達スコアで 4 位であった SMT5 が、3 種の自動評価結果ではいずれも 2 位となっており、特に BLEU、NIST では、本来品質上位であるはずの RBMT3 を大きく上回るスコアが与えられている。粗ぶるいの際には、こうした翻訳方式間の有利不利を考慮しなければならない。

今回の実験結果では、いずれの自動評価結果も、RBMT と SMT それぞれの方式内における相対順位は、人手評価結果と一致した (RBMT1<RBMT2<RBMT3、SMT4>SMT5)。

⁶ 詳細は添付資料 9-1 を参照。

この傾向は、2.1.1.6 で述べた内容とも整合しており、同一の方式のエンジン間の比較であれば、自動評価は人手評価とある程度の相関はあると考えられる。したがって、粗ぶるいの際には、同一の翻訳方式内において他よりも顕著に低いスコアのエンジンを除外するという方法も考えられる。

9.7. エンジン性能の総合的な評価

本調査における品質評価実践は、あくまで評価手法の妥当性を検証するためのものであったため、評価対象としたエンジンの大半をほぼ初期設定で利用している。しかしながら、今後実際に機械翻訳エンジンを相対評価する場合は、各エンジンのカスタマイズ機能等を利用して、それぞれを特許文献翻訳に最適化し、実用環境に極力近づけた設定で比較評価することが望ましい。

多くの機械翻訳エンジンは、翻訳品質を利用者の用途に合わせてチューニングするためのカスタマイズ機能を備えている。各エンジンのカスタマイズ機能はそれぞれ手法が異なるが、一般的には、目的とする文書の種類に応じて翻訳知識を追加していくというものが多い。添付資料 9-2 に、いくつか具体的なカスタマイズ機能の例を示した。

こうしたカスタマイズ機能も機械翻訳エンジンの性能の一部であり、各エンジンについて、これを最大限利用した状態での出力結果を比較することで、実用時の品質に則した品質評価が行える。

なお、機械翻訳エンジンの相対比較において最も重要な観点が出力結果の翻訳品質であることは言うまでもないが、これに加え、仮に評価時点で十分に活用できなかったとしても、こうしたカスタマイズ機能の幅広さや利用の容易さ、さらには各種の支援機能等も、将来的な翻訳品質の改善可能性という意味においては、エンジンの潜在能力のひとつとして考慮すべきである。例えば、出力結果において検出された翻訳ミスがこれらの機能により改善可能な性質のものであるか、またその際の対処の難易度がどの程度であるか等を分析することが望ましい。

さらに、エンジンの優劣には、インタフェースの利便性や価格、ユーザ側でコントロール可能な要素の範囲など、翻訳品質以外にも多数の観点が存在する。機械翻訳エンジンの相対評価にあたっては、翻訳品質だけでなく、これらの要素も観点に加えるべきである。

9.8. 検索システムを用いた評価

「特許庁手法」では、機械翻訳結果の主要な用途のひとつとして「検索」用途を想定しており、また、その実用環境として、「特許庁内で使われている既存システムのコンテンツ」としての使用を想定している（4.1.2.）。

「特許庁手法」では、1回の品質評価結果で「検索」用途と、これと同等に主要な用途である「粗読」用途・「精読」用途のそれぞれへの有用性を判断できるよう、いくつかの評価観点を設け、各観点の評価結果を用途に応じて重み付けする方式とした。

その結果、「検索」用途の有用性の判断は、主に「重要技術用語の翻訳精度の評価」において、適訳率がどの程度かであるかで測るという、間接的な形となっている。

しかしながら、本来「検索」用途の有用性を直接評価するのであれば、実用環境、もしくはそれに準ずる環境において、実際に機械翻訳結果を検索用データとして試用し、実際の検索における再現率（recall）と適合率（precision）等にて有用性を判断するほうが、より実用に則していることは言うまでもない。

現実には、このような評価環境を整えるのは容易でないことは事実だが、NTCIR-7の特許翻訳タスクでは、「特許庁手法」の前提とは若干異なる（検索キーワードに対して機械翻訳を適用）ものの、実際の検索結果に基づいた機械翻訳エンジンの翻訳精度を評価した例がある⁷。

機械翻訳結果の翻訳ミスの中には、検索システムの検索支援機能（例えばシソーラスを利用した検索キーワード展開機能）でカバー可能なものと不可能なものがあり、実際の検索環境を用いた品質評価を行うことで、こうした点を評価結果に反映させることができる。

⁷ 詳細は添付資料 9-3 を参照。

添付資料集

- 添付資料 1-1. 特許庁における機械翻訳品質評価に関する基本的な考え方
- 添付資料 1-2. 仮設機械翻訳品質評価手法
- 添付資料 1-3. 特許庁機械翻訳品質評価手法
- 添付資料 1-4. 第 1 回委員会 議事内容及び主な提言・指摘事項
- 添付資料 1-5. 第 2 回委員会 議事内容及び主な提言・指摘事項
- 添付資料 3-1. 「英日機械翻訳の現状分析」 サンプル文献一覧
- 添付資料 3-2. 「英日機械翻訳の現状分析」 文献別重要技術用語一覧
- 添付資料 5-1. 英日テストセット データサンプル
- 添付資料 6-1. 品質評価作業用ファイル データサンプル
- 添付資料 7-1. 特許庁機械翻訳評価手法 評価作業手順書
- 添付資料 7-2. 「流暢さの評価」 の評価サンプル
- 添付資料 7-3. 「内容の伝達レベルの評価」 の評価サンプル
- 添付資料 7-4. 「重要技術用語の翻訳精度の評価」 の評価サンプル
- 添付資料 7-5. 「チェックリストに基づく評価（言語共通項目）」 詳細説明
- 添付資料 7-6. 「チェックリストに基づく評価（英日項目）」 詳細説明
- 添付資料 7-7. 品質評価結果データ データサンプル
- 添付資料 9-1. 自動評価の粗ぶるいへの利用適否実験
- 添付資料 9-2. 機械翻訳エンジンにおける一般的なカスタマイズ機能例
- 添付資料 9-3. NTCIR 特許翻訳タスクにおける言語横断特許検索による Extrinsic 評価実験に関する調査

特許庁における機械翻訳品質評価に関する基本的な考え方

1. 機械翻訳の対象

機械翻訳の対象は、特許文献である。

- ・会話文等と比較して、文法の間違いや構造の崩れた文は少ない。
- ・元々、技術的内容を正確に伝達することを目的に書かれているため、比較的丁寧な文章となっている可能性が高いが、あまり流暢でない場合も多い。
- ・技術的専門性の高い用語や、特許文献特有の表現（請求項の記載等）が多く、通常の技術文献より文章の難易度は高い。

2. 機械翻訳結果の用途

用途①：先行技術調査の対象となる検索用データとしての利用。機械翻訳結果にはテキスト検索が行われ、検索時にヒットしたもののみが照会の対象となる。

<ポイント>

- ・先行技術調査の対象となる検索用データとして利用されることを考慮すれば、技術用語の適切な翻訳や、未知語対応が重要な観点となる。
- ・「技術用語が不適切な訳となって検索時にヒットしない」という場合と、「技術用語に様々な翻訳結果があてられ、どれが正しい翻訳結果か分からない」という場合とを比較したとき、検索後に調査者が内容確認を行うことを考慮すれば、後者の方が相対的に望ましい（前者の場合、検索漏れにつながり、調査者の目に触れることがなくなるため）

用途②：照会（粗読）用データとしての利用。調査者は、先行技術のあたりをつけるために機械翻訳データを粗読するが、引用発明としての認定や最終的な権利範囲の判断等を行う際には、原文又は人手翻訳文の内容の確認を行うことが前提。

<ポイント>

- ・調査者が読んで、精読の必要性が判断できる程度に、大まかな内容が把握できることが重要。
- ・照会用にも利用されることを考慮すれば、日本語としての流暢さも品質評価の観点となるが、粗読後の最終的な判断には機械翻訳結果は利用されないため、全評価観点中での流暢さの優先度は低い。
- ・特許文献の照会（粗読）においては、前後の文章、該当案件の図面等も参酌して内容理解を行うことが前提となるため、文脈、図面等と合わせて概要理解が可能な文については、高い評価を付与する余地がある。

3. 品質評価の目的・利用方法

品質評価には、以下の三つの目的・利用方法がある。

- 相対評価
複数のシステム間の翻訳精度比較、特定のシステムにおける翻訳品質の定点観測（経時的変化の確認）等にご利用
- 絶対評価
マッチング（※）による用途に適した機械翻訳システムの採用の検討、所定の品質水準の機械翻訳システムの用途判断、今後の対応策検討等にご利用
- FB 利用
機械翻訳システムの具体的な弱点を把握することで品質向上等にご利用

※評価点と機械翻訳活用用途のマッチングについて

評価結果の有効な活用のためには、この点数であればこの用途にこの程度貢献する、あるいはこの用途であればこの程度の点数が必要となる、といったマッチングを確立しておく必要がある。

- （例）品質 A（スコア 5）：権利侵害の存否に利用できる（人手翻訳レベル）
- 品質 B（スコア 4）：実態審査での精読に利用できる
- 品質 C（スコア 3）：検索及び粗読に利用できる（技術キーワードにより検索が可能であり、ヒットした文献を読んで精読の必要性が判断できる程度に発明の大まかな内容が把握できる）
- 品質 D（スコア 2）：検索又は粗読に利用できる
- 品質 E（スコア 1）：特許調査用機械翻訳データとしては利用不可能

4. 評価観点

- （1）技術用語の適切な翻訳（用途①と用途②の観点）
 - （2）構文、文法の適切な翻訳（用途①と用途②の観点）
- 解析の失敗（主語、動詞、目的語の関係、係り受け関係、原因・結果関係を表す接続語、省略主語への対応）、形式的に判断できる最低限の一般的な品質事項（未知語の数、処理失敗の文の数、訳抜け、不要語等）
- （3）文全体としての理解可能度（用途②の観点）
- （2. との関係で、用途①は、主に評価観点（1）（2）で担保される。用途②については、発明の内容把握に必要とされる、発明の構成、課題、機能、効果等の理解において、評価観点（1）（2）は重要であるものの、それだけでは評価しきれない文全体としての分かりやすさ、読みやすさ等が存在するため、評価観点（3）も必要である。）

5. 客観性の高い評価

評価の根拠について透明性が高く、公正・適正と考えられる評価を行うためには、基本的に「客観性の高い評価方法」を採用する必要がある。また、継続的な評価や大量の評価を人手により行う場合にも、客観性の確保が必要である。

ただし、文全体としての理解可能度は主観評価に頼らざるをえないため、客観的評価と主観的評価の両者を組み合わせることになる。

4. (3) 文全体としての理解可能度（内容伝達レベル）は、4. (1)、4. (2) を含む全体の評価につながるものであるから、メイン項目として評価を実施する（文全体としての理解可能度は、主観的な評価とならざるを得ないが、サンプル評価による担当者間のレベル合わせ、複数人による重複チェック等により客観性を担保する）が、評価における客観性の担保、及び、検索性データとしての翻訳品質評価の観点から、(A) 誤訳、未知語等の語数カウント、(B) 誤りやすいポイントを列挙したチェックリストを列挙したチェックしたチェックリストへのチェックを行う。この (A)、(B) に関しては、ポイント数による客観的評価を行うと共に、誤訳の傾向を把握する効果もあり、FB 利用にも役立つものであると考えられる。

なお、基本的には人手による品質評価を前提として手法を検討するが、客観性の担保という観点では、児童品質評価の活用も有効である可能性がある（7. 参照）。

6. 評価者の前提

特許庁及び翻訳品質評価者の人数・時間におけるリソースを考慮し、日本人による評価を前提とし、原文読解の負担も踏まえ、日本語の人手翻訳文又は日本語パテントファミリー文献との比較を前提とする。

7. 粗ぶるいについて

機械翻訳品質評価において自動品質評価の活用も有効である可能性があるため、人手による品質評価に先立ち、評価対象の言語等に応じて、機械翻訳結果の自動品質評価を行うことによって、機械翻訳結果の「粗ぶるい」を行うことが可能であるか検討する。

※相対評価に利用する場合は、自動品質評価結果を互いに比較して、自動品質評価が悪い機械翻訳システムを排除する。

※絶対評価に利用する場合は、かなり低めの閾値を定めて、閾値以下の評価がされた機械翻訳システムは、自動的に評価結果を最低ランクとする。

※技術用語の対応の的確性を重視した自動評価手法として、BLEU、NIST 等が、技術用語の出現の順序性を加味した自動評価手法として、IMPACT、RIBES 等が存在する。

仮設機械翻訳品質評価手法

➤ 事前準備

1. パテントファミリー関係等を利用して、技術的内容が対応する文献対から技術的内容が対応する文対を抽出することにより、可能な限り、技術分野（IPC）、文献種別、原文の長さ、記載箇所、出願時期、出願人等に偏りなく、機械翻訳対象文（翻訳元の言語）と参照用翻訳文（翻訳先の言語）とのセンテンス単位のペアを作成し、それを集積したテストセットを用意する（言語対ごとに 1000 文程度）。

※機械翻訳対象文に問題（助詞の欠落や誤字・脱字等）があるものは、機械翻訳の適切な評価ができないため、テストセットからは除外する。

※パテントファミリー関係の文献であっても、常に対応する文章となっているとは限らないため、テストセットの用意にあたっては、機械翻訳対象文と参照用翻訳文との対応関係についての人手によるチェックを実施する。

※テストセット作成時に、特許文献機械翻訳において重要と想定される技術用語及びチェックリストに該当する対象文を特定しておく。

2. テストセットから評価対象件数の機械翻訳対象文を抽出する（評価対象件数は 200 文程度）。機械翻訳対象文の抽出は、テストセットの作成と同様に、技術分野等について偏りないように行うこと。

3. 機械翻訳対象文について、品質評価の対象となる機械翻訳システムによって、機械翻訳を実施する。

4. 機械翻訳結果の評価に際し、評価を簡便に実施するために必要なツール（人手評価補助ツール）の作成をする。本ツールは、機械翻訳対象文、参照用翻訳文、及び機械翻訳対象文の機械翻訳結果からなる品質評価対象データをインポートし、評価者にシート形式で提示するとともに、以下のような機能を有する。

- ・機械翻訳結果に 5 段階評価を行う場合、ワンクリック等の簡潔な操作で任意の段階の評価を行うことができる。
- ・機械翻訳対象文や機械翻訳結果について、評価者が、備忘録や評価根拠のコメントを付すことができる。
- ・機械翻訳対象文について、評価者の指定範囲を自動的に辞書引きすることができる。

➤ 人手による品質評価の実施

1. 評価の観点及び基準

(1) 内容の伝達レベルの評価

＜評価基準に基づき、1－5点の評価を行う＞

5：すべての重要情報は正確に伝達されている。

4：ほとんどの重要情報は正確に伝達されている。

3：翻訳は、完全に理解できるというわけではないが、いくつかの重要情報は正確に伝達されており、正確でない情報についても大まかな内容理解が可能である。文法的、文体的に不自然な部分もあるが、内容は十分に反映されている。

2：いくつかの重要情報は正確に伝達されており、正確でない情報についても、他の文章や図面の参酌、十分な検討の時間等が与えられれば、大まかな内容を理解できる可能性がある。

1：翻訳は全く理解できない。正確に伝達されている情報はほとんどない。

＜評価対象件数について、複数人（3人以上）による重複評価を行う＞

＜評価は評価対象となる各案件の属する各技術分野に精通した評価者が行う＞

＜評価に先立って、サンプルデータ及びサンプル評価の意識合わせを行う＞

＜複数人による評価結果の平均値を算出し、さらに評価対象件数の平均値を算出する＞

＜平均値を、内容伝達レベルの評価値とする＞

(2) 技術用語に関する誤訳の個数

・ 誤訳の語数のカウント

対象文献の技術分野を鑑み、特許文献において重要と想定される技術用語の訳が、人手翻訳に対して技術的に同義な範囲を外れているもの（未知語、訳漏れを含む）の数をカウントする。

＜技術用語のチェックは、評価対象となる各案件の属する各技術分野に精通した評価者が行う＞

(3) チェックリストに基づく評価

＜評価は評価対象となる各案件の属する各技術分野に精通した評価者が行う＞

機械翻訳特有の課題を対象言語に応じて洗い出したチェックリストを作成し、チェックリストに該当する翻訳ミスの個数をカウントする。

＜課題の例＞

構文解析に関する誤り

- ・主語、動詞、目的語（他動詞の場合）の組み合わせが誤っている。
- ・主語、動詞、目的語（他動詞の場合）の対応が把握できない。
- ・動詞が複数ある場合に、副詞（句）の係り受け先が誤っている。
- ・名詞が複数ある場合に、形容詞（句）の係り受け先が誤っている。
- ・重複した係り受け（複数の形容詞（句）が1つの名詞に係り受けする場合、及び、複数の副詞（句）が1つの動詞に係り受けする場合）の訳出を誤っている。
- ・並列構造の取り違い

文法事項に関する誤り

- ・品詞を取り違えている。
- ・時制が誤っている。
- ・修飾語と被修飾語の位置関係が不適切。

語句、フレーズに関する誤り

- ・専門用語に対して一般的な訳語となっている。
- ・異なる技術分野の用語に訳されている。
- ・多義語が正しく訳されていない。
- ・未知語が含まれている。
- ・訳抜けがある。
- ・特許文献に定型的な表現（～を特徴とする〇〇、請求項〇記載の～、等の表現）に対応できていない。
- ・発明の名称について体言止めができていない。
- ・出願人名が正しく訳されていない。

記号等の処理に関する誤り

- ・数式、化学式の部分の認識誤り（途中で分断して訳出してしまっている。）
- ・箇条書き表現（";", "(", "(a)", "a.", "(1)", "1.", 等）が適切に処理されていない。
- ・本体部分とカッコ書き部分の位置関係が不適切。
- ・数字部分と単位部分の位置関係が不適切。

翻訳言語に応じた項目（中日）

- ・分離前置詞（「在～上」で「～の上に、～に関して」、等）の翻訳誤り
- ・4文字熟語の翻訳誤り
- ・目的語が前置されている場合（目的語の強調のために、SOVとされる場合がある。）の翻訳誤り。
- ・表記の一部を重ねて表現する重ね型に対処できていない。（例：「散歩（散歩する）」→「散散歩（ちょっと散歩する）」）
- ・外来語の翻訳誤り。

翻訳言語に応じた項目（韓日）

- ・ 外来語の翻訳誤り。
- ・ 同音異義語の翻訳誤り。

翻訳言語に応じた項目（日中）

- ・ 日本語の「～した」について、完了の意味でないものについても完了の意味（「了」）で訳している。
- ・ 日本語の「～される」を誤って受身に訳している。
- ・ 日本語の「～させる」を使役で訳して不自然になっている。
- ・ 原文にはあるが、訳さない方が良い場合に対応できているか（例：日本語の「～するものである」を機械翻訳すると、「もの」に対応する中国語が訳出される場合があるが、無理に訳さない方がよい訳文になる）。

翻訳言語に応じた項目（日英）

- ・ 造語（接尾語を付す：「～化」→**lightweight-ized**、ハイフンにより単語を連結：**coal-for-coke-making hopper**）が特に目立つ問題点として挙げられている
- ・ 文法の誤り
- ①時制の適・不適（関係代名詞節が **was** ないし **be** 動詞の抜けた過去分詞で始まる場合が多い。**may be, will, should** の使用規則が不明。）
- ②冠詞の用法の適・不適（冠詞の使用規則が不明。**a** と **the** の使用上の逆転が多い。冠詞が必要であるにもかかわらず付されていない。）
- ③前置詞の用法の適・不適（画一的に **by** あるいは **in** を使用している場合が多い。）
- ④関係代名詞の用法の適・不適（関係代名詞を積極的に使用しているが、多くの場合過去分詞で始まる従属節で代用可能）
- ⑤小文字・大文字の用法の適・不適（文中で大文字の動詞で始まる文章が存在する）
- ⑥命令形の用法の適・不適（原文が命令形でないにもかかわらず、命令形で表現されている翻訳が存在）
- ⑦所有格の用法の適・不適
- ⑧単数形・複数形の用法の適・不適（単数形と複数形が同形の名詞が複数形の場合、さらに複数語尾が付されている。複数形であるべき表現のものが単数形表現となっている場合がかなり存在）
- ・ 代名詞が指し示す具体的な名詞が存在するか（【課題】と特に【解決手段】において、長文の場合、代名詞 **it** の多用が見られ、しかも代名詞 **it** の指し示す具体的な名詞が不明確な場合が多い）
- ・ ピリオドがつくべき位置に付けられているか
- ・ 複合動詞（例：固定指示する→**fix and support**）への対処。
- ・ 複合動詞（形態素的には複数語から成り立っているが翻訳の観点からは1語の助詞として扱う方が文全体の係り受けとしては正しく解釈できるし、訳も適切な訳を出せるケース 例：小動物を餌としている肉食の魚→**carnivorous fish eating the small animal**）

への対処。

- ・タイトル翻訳などで不要な冠詞を訳出しないような設定を適用。

2. 相対評価方法

1. (1)～(3)の評価結果をそれぞれスコア化し、観点毎に、複数の機械翻訳システムの中で、相対評価を行う。

3. 絶対評価方法

(0) 絶対評価スコアと用途とのマッチングは、以下のとおりとする。

スコア 5：権利認定、引用発明認定に適した翻訳品質

スコア 4：実体審査、先行技術調査における精読に適した翻訳品質

スコア 3：先行技術調査及び文献紹介時の粗読及び検索データに適した翻訳品質

スコア 2：粗読又は検索データに適した翻訳品質

スコア 1：特許調査に際しては全く利用できない翻訳品質

(1) 内容の伝達レベルの評価（平均値）に基づいて、仮スコアを定める。

内容伝達：4.5～5.0 → 仮スコア：5

内容伝達：3.5～4.5 → 仮スコア：4

内容伝達：2.5～3.5 → 仮スコア：3

内容伝達：1.5～2.5 → 仮スコア：2

内容伝達：1.0～1.5 → 仮スコア：1

(2) 技術用語の翻訳について

技術用語の翻訳に関する評価結果を踏まえ、正解率が所定割合以下の場合は、スコアを1ランク下げる。

(3) チェックリストを利用した評価について

チェックリストを利用した評価結果に関して、特筆すべき問題点が存在する場合は、スコアを1ランク下げる。

また、チェックリストを利用して、翻訳ミスの傾向を分析する。

特許庁機械翻訳品質評価手法

➤ 事前準備

1. 実際の特許文献から機械翻訳対象文（翻訳元の言語）を選定し、対応する参照用翻訳文（翻訳先の言語）とのセンテンス単位のペアを作成し、評価用テストセットを用意する（言語対ごとに 1,000 文程度）。テストセットに選定する文は、可能な限り、技術分野（IPC）、文献種別、原文の長さ、記載箇所、出願時期、出願人等に偏りのないようにする。

※機械翻訳対象文は、対応する言語による対応特許の存在しない特許文献から選定し、人手翻訳にて参照用翻訳文を新規作成することが理想的であるが、テストセットの用意に使用できるリソースに制限がある場合等は、パテントファミリー関係を利用して抽出した対訳文対を用いてもよい。

※機械翻訳対象文に問題（誤字・脱字や、複数の解釈が可能な曖昧な係り受け等）があるものは、機械翻訳の適切な評価ができないため、テストセットからは除外する。

※パテントファミリー関係の文献であっても、常に対応する文章となっているとは限らないため、テストセットの用意にあたっては、機械翻訳対象文と参照用翻訳文との対応関係についての人手によるチェックを実施し、対応関係に問題がある場合はこれを除外する。参照用翻訳文を修正することで正しい対応が得られる場合は、修正を施したうえで採用してもよい。なお、機械翻訳対象文の修正は認めない。

※テストセット作成時に、各文に含まれる重要技術用語をあらかじめ選定し、評価作業時に提示するためにデータ上に記憶しておく。さらに、チェックリストの【任意項目】や【翻訳言語別任意項目】等の検証に適した文についても、あらかじめこれを特定し、各項目に該当する文がテストセットに一定数含まれていることを確認する。

2. テストセットから評価対象件数の機械翻訳対象文を抽出する（評価対象件数は 200 文程度）。機械翻訳対象文の抽出は、テストセットの作成と同様、技術分野等について偏りないようを行うことを原則とする。ただし、評価の目的や評価作業のリソースに配慮した調整は適宜行ってよい。
3. 機械翻訳対象文について、品質評価の対象となる機械翻訳システムによって、機械翻訳を実施する。
4. 機械翻訳結果の評価に際し、評価を簡便に実施するために必要なツール（人手評価補助ツール）の作成をする。本ツールは、機械翻訳対象文、参照用人手翻訳、及び機械翻訳対象文の機械翻訳結果からなる品質評価対象データをインポートし、評価者にシート形式で提示するとともに、以下のような機能を有する。
 - ・機械翻訳結果に 5 段階評価を行う場合、ワンクリック等の簡潔な操作で任意の段階の評価を行うことができる。
 - ・機械翻訳対象文や機械翻訳結果について、評価者が、備忘録や評価根拠のコメントを付

すことができる。

- ・機械翻訳対象文について、評価者の指定範囲を自動的に辞書引きすることができる。

➤ 人手による品質評価の実施

1. 評価の観点及び基準

(1) 内容の伝達レベルの評価

機械翻訳結果が、原文の内容をどの程度正確に伝達しているかを、人手翻訳の内容に照らして以下の 5 段階の評価基準で主観的に評価する。

- 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%)

※「重要情報」とは、原文に含まれる全ての実質情報（技術的要素とその相互関係）を指す。各重要情報の重要度と訳語の精度から、総合的な評価を行う。なお各評点に付記したパーセンテージは、重要情報が正確に伝達されている度合いの大体の目安である。

※機械翻訳結果の「流暢さ」については「(3)流暢さの評価」で評価するため、ここでは考慮しない。

＜評価対象件数について、複数人（3人以上）による重複評価を行う＞

＜評価は評価対象となる各案件の属する各技術分野に精通した評価者が行う。1つの機械翻訳対象文に対し複数の機械翻訳結果を評価する場合は、その全てを同一の評価者が同時に評価する＞

＜評価に先立って、サンプルデータ及びサンプル評価の意識合わせを行う。特に、「絶対評価」を目的とする場合には、「相対評価」の場合と比べて入念な意識合わせを行う。具体的には、各評価者が、正解データが存在する訓練用評価文 30 文程度にて評価を試行し、評価結果と正解データを照合して自らの評価水準を意図されたレベルに修正したうえで、実際の評価対象データから選定した 30 文程度について、各評価者が評価結果を持ち寄り、統一の評価結果を話し合いにより定め、評価作業管理者の承認を得た後、これを目安として残りの評価対象文の評価を行う。なおサンプル評価結果は、次回以降のリファレンスとして保存・活用する＞

(2) 重要技術用語の翻訳精度の評価

・重要技術用語の訳語の評価

対象文の技術分野を鑑み、各機械翻訳対象文に対してあらかじめ選定された重要技術用語（1語～3語）に対応する機械翻訳結果中の訳語を、以下の基準で評価する。

A（適訳）：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）：技術用語として一般的に用いられる訳語ではないが、意味はおおむね正しい。

C（誤訳）：誤訳である。

D（不訳）：未知語、訳漏れである。

※同一の重要技術用語が1文において複数回出現し、異なる訳語に訳されている場合は、全ての訳語を上記基準で評価し、最も高い評価点（A＞D）のみを入力する。

※機械翻訳対象文が重要技術用語を含まない場合は、評価対象外とする。

・正解率、誤訳率の算出

各重要技術用語の評価点を集計し、評価点 A～D それぞれの比率を算出する。さらに用途に応じた重みづけを行い、正解率、誤訳率として評価値に利用する。

＜技術用語のチェックは、評価対象となる各案件の属する各技術分野に精通した評価者が行う。1つの機械翻訳対象文に対し複数の機械翻訳結果を評価する場合は、その全てを同一の評価者が同時に評価する＞

(3) 流暢さの評価

機械翻訳結果を独立した文として見た際の、文としての自然さ、読みやすさのみを以下の5段階の評価基準で主観的に評価する。

5：文意が明解で、人間が書いた文に近い。

4：文として不自然な箇所を若干含むが、文意は明解である。

3：文として不自然な箇所があり、文意がわかりにくい。

2：文法規則に反する表現をかなり含む。文意がわからない。

1：文として成立していない。

※本評価は、翻訳結果のみを独立した日本語文として扱う。「文意」とは、あくまで翻訳結果において読み取れる文意を指し、原文や人手翻訳の内容と一致しているかどうかは考慮しない。

＜1つの機械翻訳対象文に対し同一の評価者が「流暢さ」の評価と他の観点からの評価の双方を担当する場合、最初に「流暢さの評価」を、原文や人手翻訳を参照せずに行う＞

＜1つの機械翻訳対象文に対し複数の機械翻訳結果を評価する場合は、その全てを同一の評価者が同時に評価する＞

(4) チェックリストに基づく評価

選定した機械評価対象文の機械翻訳結果に含まれる翻訳ミスの個数を、以下に示すチェックリストに基づきカウントする。

チェックリストは、以下に示す【必須項目】をベースとし、必要に応じて【任意項目】及び【翻訳言語別任意項目】から任意に選択した項目をこれに追加する。

評価対象文数は任意に設定してよい。

【必須項目】

機械翻訳結果に含まれる全ての翻訳ミスを、下記 A～G の該当する項目 1 箇所にかウントする。1 つの翻訳ミスが下記項目の複数に該当する場合は、上位の項目のみにカウントする。

- A. 語の欠落
- B. 不要語の混入
- C. 未知語
- D. 品詞の誤り
- E. 係り受け誤り
- F. 語の誤訳
- G. その他の翻訳ミス

【任意項目】

【必須項目】にかウントした翻訳ミスについて、特に下記項目にあてはまる場合はここにもカウントする。1 つの翻訳ミスが下記項目の複数に該当する場合は、該当する全ての項目にかウントする。

- Z-00. 訳語のゆれ
- Z-01. 不適切な文の分断
- Z-02. 主語・述語・目的語（他動詞の場合）の相互関係に関する翻訳ミス
- Z-03. 並列表現に関する翻訳ミス
- Z-04. 発明の名称など、名詞句を文として扱った翻訳ミス
- Z-05. 特許特有の用語や表現に関する翻訳ミス
- Z-06. 特許特有の記号や文献番号表記、付番に関する翻訳ミス
- Z-07. 箇条書き表現に関する翻訳ミス
- Z-08. 原文のカッコ書きに起因する翻訳ミス
- Z-09. 数式、化学式に関する翻訳ミス
- Z-10. 数字と単位に関する翻訳ミス
- Z-11. 数値範囲に関する翻訳ミス

※【任意項目】は、評価の目的や作業負担等を鑑み、必要な項目のみを使用する。

[チェック手順]

- ・ 機械翻訳結果に含まれる全ての翻訳ミスを【必須項目】に分類し、個数をカウントする。
- ・ 翻訳ミスのカウントは、語単位（複合語や慣用表現は一語と見なす）、係り受け関係単位でのカウントを原則とする。重要技術用語として評価した語も再度評価対象とする。
- ・ 1 個の翻訳ミスは【必須項目】中の 1 項目のみにカウントする。翻訳ミスが複数の項目に該当する場合は、順序が上位の項目のみにカウントする。
- ・ 1 語に対して複数の異なる翻訳ミスが生じている場合は、個別にカウントする。
- ・ 各翻訳ミスは、【必須項目】にカウントした後、【任意項目】の各項目と照合し、該当する全ての項目にもカウントする。
- ・ 日本語として読みにくいが意味的・文法的に誤解を与える懸念の少ない語や語順、文構造は、「流暢さ」の範疇であり、ここでは翻訳ミスと見なさない。

＜評価は、機械翻訳対象文に用いられる言語の読解力を備え、かつ評価対象となる各案件の属する各技術分野に精通した評価者が行う。1 つの機械翻訳対象文に対し複数の機械翻訳結果を評価する場合は、その全てを同一の評価者が同時に評価する＞

＜評価は評価対象となる各案件の属する各技術分野に精通した評価者が行う＞

【翻訳言語別任意項目】

【必須項目】及び【任意項目】にカウントした翻訳ミスについて、特に下記項目にあてはまる場合はここにもカウントする。1 つの翻訳ミスが下記項目の複数の項目に該当する場合は、該当する全ての項目にカウントする。

※【翻訳言語別任意項目】は、評価の目的や作業負担等を鑑み、必要な項目のみを使用する。

[英日]

- EJ-01. 文中のピリオドを適切に処理していない
- EJ-02. 「:」や「;」を用いた文を適切に処理していない
- EJ-03. 重要な単数・複数情報が維持されていない
- EJ-04. 単数形・複数形併記表現を適切に処理していない
- EJ-05. 序数詞表現が正しく解釈されていない
- EJ-06. 慣用句の中間への語の挿入を適切に処理していない
- EJ-07. 倒置文を適切に翻訳していない
- EJ-08. 名詞と代名詞の関係を適切に表現できていない
- EJ-09. 慣用表現とみなすべき場合／そうでない場合を適切に翻訳できていない
- EJ-10. 前置詞に対し適切な訳語を選択できていない
- EJ-11. 接頭辞、接尾辞を伴う語を適切に処理していない
- EJ-12. スラッシュを用いた並列表現を適切に処理していない
- EJ-13. therein, therefor, wherein, 等を適切に処理していない

- EJ-14. 代名詞 **same**、**the same** を適切に処理していない (【発明の名称】)
EJ-15. 小文字を含む大文字表記を適切に処理していない (【発明の名称】)
EJ-16. クレーム引用に関する表現を適切に翻訳していない (【特許請求の範囲】)

2. 相対評価方法

「1. 評価の観点及び基準」(1)～(3)の各観点の評価結果をそれぞれスコア化し、複数の機械翻訳システム間、もしくは特定のシステムにおける複数の機械翻訳結果間で、相対評価を行う。

※「(4) チェックリストに基づく評価」の評価結果は、相対評価には用いない。

各評価結果のスコア化は、以下の手順を基本とする。

(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「内容伝達レベルの評価」の評価値とする。

(2)「重要技術用語の翻訳精度の評価」結果のスコア化

- ①評価対象とした「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～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」総数
 - ・誤訳率 = 評価 C (誤訳) の語数 ÷ 「重要技術用語」総数
 - ・不訳率 = 評価 D (不訳) の語数 ÷ 「重要技術用語」総数
- ③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。

(3)「流暢さの評価」結果のスコア化

- ①評価結果を文長別に集計し、これを特許文献の文長比率に基づき補正して、「流暢さの評価」のスコアとする。

(4) 総合評価

上記(1)～(3)の評価値を、用途等を考慮して適切に重み付けし、総合評価を行う。

3. 絶対評価方法

(0) 絶対評価スコアと用途とのマッチングは、以下のとおりとする。

スコア 5：権利認定、引用発明認定に適した翻訳品質

スコア 4：実体審査、先行技術調査における精読に適した翻訳品質

スコア 3：先行技術調査及び文献照会時の粗読及び検索データに適した翻訳品質

スコア 2：粗読又は検索データに適した翻訳品質

スコア 1：特許調査に際しては全く利用できない翻訳品質

(1) 絶対評価スコアを、「内容の伝達レベルの評価」のスコア及び「重要技術用語の翻訳精度の評価」の適訳率に基づき定める。

内容伝達 適訳率	1.0～3.0	3.0～3.5	3.5～4.5	4.5～5.0
60%以上	2	3	4	5
60%未満	1	2		

(2) 「流暢さの評価」において、特筆すべき問題点が存在する場合は、その内容に応じてスコアを減点する。

4. 主要項目と補助項目

前記 1. (1)～(4)の各評価項目を、以下のとおり「主要項目」と「補助項目」とに区分し、評価作業のリソース（人数、時間等）に制限があり、全ての評価項目を実施することが困難な場合には、「主要項目」のみを実施するものとする。

主要項目：(1) 内容の伝達レベルの評価

(2) 重要技術用語の翻訳精度の評価

補助項目：(3) 流暢さの評価

(4) チェックリストに基づく評価

また、主要項目である「(1) 内容の伝達レベルの評価」において、1 文に対し 3 名以上の評価者を用意することが不可能な場合は、評価者の人数を削減してこれを実施することも可とする。

第1回委員会 議事内容及び主な提言・指摘事項

I. 日時・会場

日時：平成25年12月6日（金）15：30－18：30

会場：特許庁 特別会議室

目的：「新設機械翻訳品質評価手法（新設手法）」の妥当性について検討する。

II. 議題

1. 新設機械翻訳品質評価手法の説明

1-1. 前提条件および＜事前準備＞部分

1-2. ＜人手による品質評価の実施＞部分

1-2-1. 評価の観点及び基準

- (1) 内容の伝達レベルの評価
- (2) 重要技術用語の翻訳精度の評価
- (3) チェックリストに基づく評価
- (4) 流暢さの評価

1-2-2. 相対評価手法／絶対評価手法／主要項目と補助項目

2. その他の課題、検討事項

III. 主な提言・指摘事項

1-1. 前提条件および＜事前準備＞部分

「新設手法」作成の際の前提条件と、同手法の＜事前準備＞部分（テストセットの準備等の定義部分）の内容について説明した。

[主な提言・指摘事項]

- ・ テストセットには対応特許から抽出した文を使用することになっているが、そうすると統計方式の翻訳システム（SMT）の学習データに含まれる可能性がある。その場合、これはフェアな評価とはいえなくなる。従って、テストセットの母集団は統計翻訳の学習データとして使えないようなデータ、理想的には人手により翻訳したデータを用意すべきである。
- ・ テストセットに選定する文は、特許文献のどの部分から採取するのか。核心的な技術的事項の部分から採取すべきではないか。また分野によっても傾向が異なる。
- ・ テストセットの文数200文の妥当性については、NTCIRの経験から考えると、200文

あれば、ある程度一般化された評価になる感触があるが、評価がしっかりとできているかどうかを検証する必要がある。100 文×2 に分割して双方矛盾のないことを検証すればよい。

- ・ 特許文献の機械翻訳結果の用途として「検索」と「粗読」用途とが想定されているのに対し、テストセットでは各記載箇所から機械翻訳で誤りやすい様々なタイプの文を選定する方針とのことだが、「検索」や「粗読」用途において重要な文に、それほど多様なバリエーションが存在するのか。もし、こうした用途のために絶対に翻訳しなければならない文のパターンが絞れるようであれば、それによって評価対象文の数を減らすこともできるかも知れない。いずれにせよ、「検索」・「照会」用途を考慮したうえで、そのために扱うべき文のパターンがどのようなものであり、件数はどの程度必要かということがダイレクトに示されれば妥当性を検証しやすくなる。
- ・ テストセットに特許文の典型的なものを選ぶこと自体は良いが、あまり簡単なものでは意味が無いと考える。できれば機械翻訳が誤りやすい表現、例えば数値範囲等が文中に含まれると翻訳精度が急激に低下するので、こうした性質の表現をテストセットに含めると有用であると考ええる。

1-2-1. 評価の観点及び基準 「(1) 内容の伝達レベルの評価」

「新設手法」の第 1 の評価観点である「(1) 内容の伝達レベルの評価」部分の定義内容について、5 段階の主観的評価であり、文中の全ての「重要情報」を対象とすること、各評価にパーセンテージを併記したこと、複数名による重複評価であること等が事務局より説明された。

[主な提言・指摘事項]

- ・ 直感的、主観的評価であるということが確実に評価者に伝わるようにするには、説明の冒頭でそれを明示すべきである。重要情報の定義は必要であるが、あくまで補足説明の位置づけにしたほうが誤解がない。
- ・ 誰が評価してもなるべく同じような評価できることが重要である。付記されたパーセンテージが何を意味しているのか詳しく説明したほうがよい。現状では、パーセントが何を何で割っているのかがよくわからないし、数え方もワードレベル、句レベルで数えるのかによって同じ翻訳結果であっても数字は変わってくる。同じ翻訳結果でも、単語レベルで考えれば 90%、一方、節レベルで考えれば 0%となる場合が考えられる。ただし、このようにして機械的に決めるのではなく、このパーセンテージは各評価グレードの目安として考えて、最終的に主観評価で決めるのが良い。
- ・ 評価サンプルをいくつか提示するのが良いかもしれない。評価者間の意識合わせができる形であればよい。評価者に事前に評価を試行してもらい、フィードバックを通じて

意識合わせをすることも大切である。

1-2-1. 評価の観点及び基準 「(2) 技術用語の翻訳精度の評価」

「新設手法」の第2の評価観点である「(2) 技術用語の翻訳精度の評価」部分の定義内容について、適訳語以外は一律誤訳とみなす「仮設手法」の評価基準を改善し、可訳語（適訳語ではないが意味は分かる語）という評点を追加したこと、検索用途の場合は「適訳語」のみ、粗読用途の場合は「適訳語＋可訳語」を有用と扱う想定であること等が事務局より説明された。

[主な提言・指摘事項]

- ・ 検索用途において可訳語では役に立たないとしているが、それは検索システムの作り方による。例えば、機械翻訳で間違いやすいことが分かっているならば、検索時にその間違いを踏まえたキーワード展開等を行うことで補える。一種のシソーラスなのかもしれないが、こうした機能の可能性を考えると評価基準も変わってくる。
- ・ 技術用語の翻訳精度の調査の目的は、検索の用途の観点で評価をしたいから設けた項目か。もしそうであれば、なぜ翻訳結果の検索精度をダイレクトに評価しないのか。
→ 必ずしも検索用途の評価に限定していない。特許文献においては技術用語の翻訳は重要な観点であるため、独立した観点として評価し、総合評価の際には都度の用途に応じて他の評価値との重み付けを変えられるようにしたほうが扱い易いであろうという考えで設定した観点である。
- 検索システムとしての評価は今回の評価とは別にやるべきだろう。検索システムを設計する側の評価も必要だ。クエリ拡張機能も、検索範囲は広がるがゴミも拾ってきてしまうので、最適性の問題等も考慮しなければならない。

1-2-1. 評価の観点及び基準 「(3) チェックリストに基づく評価」

「新設手法」の第3の評価観点である「(3) チェックリストに基づく評価」部分について、機械翻訳で生じ得る翻訳ミスを大きく4種類（A～D）に分類し、それぞれについて重大と軽微に区分することで、翻訳ミスの個数を把握しつつ、主観的評価である「内容の伝達レベル」等ともある程度整合する基準としたこと等が事務局より説明された。

[主な提言・指摘事項]

- ・ このチェックリストによる評価を何のためにやるのかというところから遡って考える必要がある。主観的に行う「内容の伝達レベルの評価」の内容が妥当であることを裏付けるために行うものなのか、それとも別の目的があつてのものか。仮設手法を見ると、並列構造の取り違えや、時制、主語述語の組み合わせが誤っている等、「内容の伝達レベルの評価」と異なり個別の観点について着目しており、別の評価と考えることもできる。

チェックリスト評価の目的によって、どのような形にするかが変わってくる。

⇒ 「内容の伝達レベルの評価」ではシステムの良否は判断できるが、なぜ良いのか悪いのか、その原因を評価結果から把握することはできない。このため、システムごとの傾向を捉えることを主眼にチェックリストを検討した。

- ・ 評価手法は機械翻訳を使う目的に依存するので、それをしっかりおさえないと良い評価指標であるかどうかかわからない。フィードバック目的であれば細かく分析するほうが良いが、このような言語学的な知識は評価者には期待できないのではないか。フィードバック目的であるとしても、正しく翻訳されない文をそのままメーカーに伝えて、メーカーで原因を究明するほうが確実ではないか。

- ・ このチェックリストは A～D 項目でそれぞれに文意への影響の大／小があって二次元になっている。システムは、原文を見てどの部分が文意にとって影響が大きいかなどは考えずに訳しており、ここが二次元であらわされている理由が分からない。

⇒ これは、主観的評価を客観的評価で裏付けるという考えから、内容の伝達レベルの評価とチェックリスト評価とで、少なくとも相対評価の順位が整合するためには、単に翻訳ミスのカウントを数えるだけでは充分でなく、重大・軽微の区別も必要なのではないかと考えたためである。

1-2-1. 評価の観点及び基準 「(4) 流暢さの評価」

「仮設手法」には「流暢さ」を評価する基準がないが、補助的な観点として必要との考えに基づき、「新設手法」にはこの評価観点が追加された。「(4) 流暢さの評価」は、50 文のサンプル評価としている。

[主な提言・指摘事項]

- ・ 50 文のサンプリングにて行うとのことだが、微妙なシステム差を判断するためには大量データの判断が必要だ。50 文で充分かどうかは疑わしい。
- ・ 機械翻訳に流暢さを求めるのは難しいと思うが、人が書いても流暢な日本語を書くのは難しい。一方で、大量の文献を素早く読む立場に立つと流暢であることは重要なファクターである。流暢さを項目に入れることに賛成するが、主観的かつ補助的な項目と考える。評価者が流暢さの評価に時間を取られないような配慮も必要なのではないか。
- ・ 流暢さの評価の際、内容の伝達レベル等、他の評価の影響を受けがちであることがわかっており、評価作業においては工夫が必要である。ALPAC レポートでは、最初に流暢さの評価をさせ、その後で正確さの評価をしている。

1-2-2. 相対評価手法／絶対評価手法／主要項目と補助項目

「新設手法」の各観点からの評価結果を用いた相対評価、絶対評価の方法が説明された。また、評価リソースに制約がある場合への対応として、「内容の伝達レベル」と「重要技術用語の翻訳精度」の2観点のみを主要項目として実施を許容することとした。

[主な提言・指摘事項]

- ・ 新設手法の場合、テストセットから一文単位の評価となるため、訳文から何を知りたいかが明らかでなければ評価しにくいのは事実だが、文の意味が分かれば役に立つので、あくまで翻訳精度の観点と割り切ればある程度は評価可能だと思う。

2. その他の課題、検討事項

「新設手法」の改善点や、「新設手法」の範囲外であるが機械翻訳結果の品質評価として考えるべき観点等について提言・指摘があった。

[主な提言・指摘事項]

- ・ 仮設手法と新設手法とでそれぞれ評価を行い、新設手法のほうが分解能が高くなりました、という検証が必要なのではないか。
- ⇒ 新設手法は、仮設手法の使いにくい点や分かりにくい部分を改善したものであり、チェックリスト以外は評価観点・基準自体は大きく変えていない。このため、両者を比較しても有意な差は出ないと思う。またチェックリストについては、仮設手法は運用基準に曖昧な部分があり、実用は難しい。
- ・ 評価対象文の文脈の中で文意をどう位置付けるか、という問題がある。今回は一文単位での評価となるが、例えば評価時に前後の文を見せるやりかたもある。
- ・ 翻訳ソフト開発の立場では、エンジンにユーザの辞書を加えるなどしてカスタマイズして良くなっていくと考えている。実際にエンジンを選定する際には、初期状態の翻訳精度・品質だけでなく、これに対しどのような対応を行えば希望する水準を達成できるか、その対応策はどのくらい簡単に行えるのか、等についても評価観点として将来的には考慮すべきである。
- ・ 翻訳システム全体の評価はヒューマンインターフェースも関係する。ただし、本調査においては初期状態のエンジンの性能を評価することになる。例えば、翻訳メモリや文パターン等を加えた後の評価は重要だが、今回そこまで範囲を広げることは難しい。
- ・ 評価をどのように実施するかということに関して、各システムの評価環境が公平になるのが大切である。内容の伝達レベルの評価においては、5段階評価の評価値そのものより、翻訳結果の優劣の比較の方が信頼性が高い。そのため、システム間の優劣が公平に評価されることが重要である。従って、対象文単位で、全てのシステムの翻訳結果が

公平な条件、つまり同一の評価者により同時に比較できる状態で評価されることが重要である。この条件を手順に明記すべきである。

- ・ チェックリストに関しては、仮設手法の内容では、具体的にこのようなミスがいくつありましたというのが分かり易いが、新設手法では、全てのミスを漏れなくカウントすることを重視した結果、大まかな種別ごとに何回のミスがあったということしか把握できなくなっている。この方針で進めるのであれば、せめて「並列表現の誤り」「主語述語の誤り」の2つだけでも、ひとつ上のレイヤーとしてカウントするようにすれば、意味の伝達レベルとは異なる観点での評価として有用となるのではないか。労力的な問題はあるかも知れないが。
- ⇒ チェックリスト評価については、なぜこの評価が必要なのか、その用途・目的を改めて整理して方針を定めることとする。現状は、チェックリストを相対評価に用いる客観的評価のひとつとして扱うという用途と、システムの傾向や弱点を把握する用途とが混ざってしまっている。従って、相対評価にチェックリストを含める必要があるのか、もしくは相対評価の対象外として、特定の課題についてのみカウントする形とすべきなのか、再度検討する。
- ・ チェックリストの英日項目を見ると、評価者に英語の知識がないと評価困難であるように見受けられる。これは、前提条件で説明された評価者の条件と矛盾するのではないか。英語であれば皆ある程度わかるだろうが、他の外国語を想定すると、このような評価項目では誰も評価できなくなることが懸念される。

以 上

第2回委員会 議事内容及び主な提言・指摘事項

I. 日時・会場

日時：平成26年2月5日（水） 15：00－17：10

会場：特許庁 特別会議室

目的：「特許庁機械翻訳品質評価手法（特許庁手法）」の実践結果から、同手法の妥当性、課題、問題点について検討するとともに、同手法を英日以外の言語対に適応する場合についても検討する。

II. 議題

1. 「特許庁手法」の内容説明
2. 「特許庁手法」の妥当性の検証
 - 2-1. 「内容の伝達レベルの評価」
 - 2-2. 「重要技術用語の評価」
 - 2-3. 「流暢さの評価」
 - 2-4. 「チェックリストに基づく評価」
 - 2-5. 「絶対評価スコア」
3. 「特許庁手法」の課題

III. 主な提言・指摘事項

1. 「特許庁手法」の内容説明

本調査にて作成した「特許庁手法」の内容について、「新設手法」からの変更点を中心に事務局より説明した。

（※事前に各委員に展開し、了承済みの内容の再説明であり、指摘・質問等はなし。）

2. 「特許庁手法」の妥当性の検証

「特許庁手法」を用いて、英日テストセットから 199 文を対象に実践した品質評価の結果と、これを用いて考察した同手法の妥当性の検証内容について、事務局より説明した。

[主な提言・指摘事項]

- ・ 相対評価において、2つのシステム間の優劣をつけるのに何文程度評価すれば充分かというのは、システム間の性能が接近していればいるほど多くの対象文が必要となるものであり、具体的にどういうシステムを相対評価したいのかに係ってくる。そこが不定である場合、何文を対象とすれば確実であるということとは言えない性質のものである。
- ・ どうしても優劣をつけたい場合、同等とみなされた両システムのスコアを計算するこ

との他に、それぞれの出力結果のうち、どちらが良いかという優劣を直接つけていくという評価手法がある。比較的实施が簡単で信頼性が高い評価手法であり、これを追加で実施するとよい。

- ・ 絶対評価スコアの検証は、本調査では 1 名の元審査官の判定に基づいており、信頼度が低い。3 名くらいでやれるとさらに信頼性が上がる。
- ・ 絶対評価スコアの検証スコアについて、内容伝達スコア及び重要技術用語の適訳率との重回帰分析を実行したが、結果とかなり違っていて結果を裏付けることができなかった。特に重要技術用語は 1 文から最大 3 語しか採っていないので、1 文単位では数字が大雑把となり、ばらつきが大きくなってしまう。そのような理由で、うまく回帰ができなくなっている。また、絶対評価を 1 人でしかやっていないことも課題となる。もう少しデータが増えれば乗る可能性はある。
- ・ 回帰をチェックしたのは内容伝達スコアと技術用語の評価だけとのことだが、流暢さの評価結果等、他のデータについても、きっちりと分析すべきである。
- ・ 絶対評価スコア 4 や 5 の評価を文単位で行うのは難しいのではないかと。審査官経験者であっても、ある程度長い文章を見ての印象でないと判断できないように思われる。また、スコア 5 については審査官個人の好みが反映してしまう懸念がある。今回の検証で、元審査官が 4 と 5 を判断する時にどういったことを考えて判断したかを把握することが重要である。
- ・ 絶対評価を行う場合は、今回の品質評価実践で行なった方式よりも入念な意識合わせを実施して、評価水準のずれを解消する必要がある。評価者の基準を合わせることは、かなり困難な作業ではあるが、できるだけ詰めた方が正確性が上がる。
- ・ 絶対評価のマッチング基準を考えるにあたり、今回は 1 名の元審査官の判定結果に基づいたが、どのようにすれば一番マッチングできるかを決めるのであれば、レファレンスを作って、そのレファレンスにおいてはこのようなマッチングがとれる、というデータを作ったらよい。そのデータというのは今回の 200 文の 5 システムの正解の内容伝達レベルというのを決め、絶対評価の値もそれぞれ決めて、マッチングの式を決めればよい。次に内容伝達レベルの評価を行う場合は、評価者の評価結果が十分同じ値になるように訓練してから、評価すればよい。
- ・ 絶対評価の評価値 4 と 5 をどうしても付ける必要があるのか。絶対評価スコア 1、2 を解決するのにエネルギーを注ぐのであれば、3 以上かどうかが分かればよいのではないかと。
- ・ 評価者の立場に立つと、サンプルが多いと確認作業の負担が増大するし、かえって混

乱する場合もあるので留意が必要である。

- ・ 「内容伝達レベルの評価」において、「重要情報」をどういう単位で何%と数えるのかをしっかりと定義する必要がある。文の否定と肯定が逆になっていると、語単位の訳語はほぼ全て合っているけれど文のレベルで見ると意味が逆となるような場合、どういう単位で評価値を考えるかというのは、評価者全員が共有しないと、評価が揺れる部分となる。
- ・ 今後の課題となるが、評価単位の問題は、小さいレベルの評価と大きいレベルで評価の次元を分けて用意すれば良いのではないか。そうすれば、評価者がそれを一元化することに悩むことは無くなる。

3. 「特許庁手法」の課題

品質評価実践を踏まえた「特許庁手法」の今後の課題、具体的にはテストセット関連、英日以外の言語対への適用可否等について議論を行なった。

〔主な提言・指摘事項〕

- ・ テストセットの再作成について、フィードバック用途にもテストセットを使うのであれば、フィードバック先である開発者にテストセットの文を開示してしまうと、その文に対しては非常によく訳せるように（ピンポイントで）チューニングされてしまうことがあるので、フィードバックの際にはそのテストセット文でなく、類似の文を開示すべきである。テストセットは問題点の発見にだけ使うのがよい。
- ・ テストセットは、パテントファミリーの利用するのではなく、ローカル文献に新たに人手翻訳をつける形とするのが良い。ローカル文献であれば、現状の SMT 技術では、原語（英日翻訳であれば英語）で記載された原文のみをチューニングに利用することは容易ではないため、評価条件が不平等となることは考えにくい。なお、「テストセットを学習データから除外する」という方法は、SMT 方式といえども商用システムでは学習データの管理が複雑化するため、ここから特定のデータだけを削除することは困難であり、実現性に乏しい。
- ・ 今回の検証結果から、「特許庁手法」は、相対評価によるシステム比較をする場合には十分に使える。評価対象文も、システムの性能が非常に近い場合を除き、200 文で妥当である。ただし、絶対評価については、まだ課題がある。今後さらに評価方法とマッチング基準を定めて行くのが良い。

以 上

「英日機械翻訳の現状分析」 サンプル文献一覧

	IPC	文献番号	タイトル
1	A01B59/043	US2007000673	Three point hitch assembly
2	A47B57/06	US2012000865	VERTICALLY ADJUSTABLE SUPPORT DEVICE FOR LAUNDRY
3	A61B5/14	EP2400885	ECG DATA DISPLAY METHOD FOR RAPID DETECTION OF MYOCARDIAL ISCHEMIA
4	B23K10/00	US2008000880	System and method for treating a coating on a substrate
5	B65D85/48	EP1737764	CONTAINER
6	B02C17/10	EP1737575	TESTING METHOD FOR BALL MILLS
7	C30B23/00	US2007000432	Method for growth of silicon carbide single crystal, silicon carbide seed crystal, and silicon carbide single crystal
8	C03B9/453	EP1737798	TRANSFERRING ELEMENT OF A PUSHER MECHANISM OF A GLASS FORMING MACHINE
9	D06F75/38	US2008000115	Iron Comprising a Thermal Screen With an Integrated Conduit
10	D02G3/44	EP2140051	ENGINEERED TEXTILE YARN
11	E06B9/56	US2008000596	Device for fixing a pulley
12	E01F15/06	EP2140068	PERIMETER SECURITY BARRIERS
13	F02M35/10	US2008000540	Flexible Hose
14	F25B41/04	EP1738119	A COOLING DEVICE AND CONTROL METHOD
15	G01F1/00	US2009000389	MULTIPLE PATH AIR MASS FLOW SENSOR ASSEMBLY
16	G09F13/16	US2009000166	Day/night safety sign using reflective internal lighting
17	G06F9/54	EP1872211	METHOD AND SYSTEM FOR EXECUTING A CONTAINER-MANAGED APPLICATION ON A PROCESSING DEVICE
18	H05B6/80	US2010000992	Microwave Steamer
19	H01L23/60	EP2269220	ESD NETWORKS FOR SOLDER BUMP INTEGRATED CIRCUITS
20	H04W36/08	EP2401878	METHOD FOR USE IN MOBILITY HAND-OFF BETWEEN BASE STATIONS

「英日機械翻訳の現状分析」文献別重要技術用語一覧

[文献#1 : US2007000673]

	重要技術用語	日本語人手翻訳	品詞
1	coupling assembly	連結組立体	
2	cross sleeve	横向きスリーブ	
3	extension	拡張	
4	three point mounting hitch	3点装着ヒッチ	
5	telescopically	入れ子状に	副詞
6	arm	アーム	
7	hook	フック	
8	spread distance	展開距離	
9	retraction	格納	
10	receive	収容する	動詞

[文献#2 : US2012000865]

	重要技術用語	日本語人手翻訳	品詞
1	end bearing	末端ベアリング	
2	laundry items	洗濯物	
3	motorized shaft	モーター駆動シャフト	
4	pulley	プーリー	
5	freely rotatable shaft	自由に回転可能なシャフト	
6	rotatably supported shaft	回転可能に支持されているシャフト	
7	remotely controlled, reversible, electric motor	遠隔制御される逆回転可能な電気モーター	
8	vertically adjustable support device	垂直方向に調節可能な支持装置	
9	horizontal surface	水平な表面	
10	drive housing	駆動筐体	

[文献#3 : EP2400885(A1)]

	重要技術用語	日本語人手翻訳	品詞
1	ST segment deviation	S T 部分偏位	
2	multi-lead ECG	多誘導 E C G	
3	zero ST deviation value	ゼロ S T 偏位値	
4	patient monitoring device	患者モニター装置	
5	polar diagram	極図式	
6	continuous ST segment monitoring	連続 S T 部分モニタリング	
7	12-lead ECG	1 2 誘導 E C G	
8	mathematical linear combination	数学的一次結合	
9	ischemic event	虚血事象	
10	ST depression	S T 下降	

[文献#4 : US2008000880]

	重要技術用語	日本語人手翻訳	品詞
1	multilayer coating	多層被膜	
2	substrate	基板	
3	thermoelectric structure	熱電構造体	
4	energy source	エネルギー源	
5	boron carbide/boron carbide system	炭化ホウ素／炭化ホウ素系	
6	physical vapor deposition	物理蒸着	
7	nanosecond pulsed laser	ナノ秒パルスレーザー	
8	coating melting temperature	被膜融解温度	
9	substrate degradation temperature	基板分解温度	
10	electrical conductivity	導電率	

[文献#5 : EP1737764]

	重要技術用語	日本語人手翻訳	品詞
1	container unit	コンテナ装置	
2	sheet item	薄板品	
3	windscreen	フロントガラス	
4	securing arrangement	固定用装置	
5	upright restraint	直立固定具	
6	lock arrangement	係止装置	
7	brake system	ブレーキシステム	
8	brake bar	ブレーキバー	
9	slide bar	スライドバー	
10	side panel	側板	

[文献#6 : EP1737575]

	重要技術用語	日本語人手翻訳	品詞
1	testing method	試験方法	
2	Bond ball mill	ボンドボールミル	
3	semiautogenous	半自生	形容詞
4	autogenous	自生	形容詞
5	grinding circuit	粉砕回路	
6	ore	鉱石	
7	grinding energy	粉砕エネルギー	
8	particle size	粒子サイズ	
9	Bond Mill Work Index	ボンドミルワークインデックス	
10	feed	原料	

[文献#7 : US2007000432]

	重要技術用語	日本語人手翻訳	品詞
1	single crystal	単結晶	
2	silicon carbide	炭化ケイ素	
3	seed crystal	種結晶	
4	crystal growth	結晶成長	
5	(0001) plane or (000-1) plane	(0 0 0 1) 面または (0 0 0 - 1) 面	その他
6	sublimation technique	昇華技術	
7	crystal defect	結晶欠陥	
8	off angle	オフ角	
9	graphite crucible	黒鉛るつぼ	
10	washing treatment	洗浄処理	

[文献#8 : EP1737798]

	重要技術用語	日本語人手翻訳	品詞
1	transferring element	移送構成要素	
2	pusher mechanism	プッシャ機構	
3	glass forming machine	ガラス成形機	
4	pressurized air	加圧空気	
5	clamping form groove	クランプ形状溝	
6	distribution channel	分配チャネル	
7	contact insert	接触インサート	
8	dead plate	デッドプレート	
9	conveyor belt	コンベアベルト	
10	retaining force	保持力	

[文献#9 : US2008000115]

	重要技術用語	日本語人手翻訳	品詞
1	conduit	導管	
2	heating body	加熱ボディー	
3	housing	ハウジング	
4	heat shield	熱シールド	
5	iron	アイロン	
6	molding	成型	
7	pump	ポンプ	
8	reservoir	リザーバ	
9	soleplate	ソールプレート	
10	steam chamber	蒸気チャンバ	

[文献#10 : EP2140051]

	重要技術用語	日本語人手翻訳	品詞
1	electrically conductive textile yarn	導電性織り糸	
2	protective garment	防護服	
3	resistance to fire	耐火性	
4	filament	フィラメント	
5	fibre	ファイバー	
6	thread	スレッド	
7	electrically conductive filament	導電性フィラメント	
8	core	コア	
9	resistance to heat	耐熱性	
10	staples of first element	第1の要素の束	その他

[文献#11 : US2008000596]

	重要技術用語	日本語人手翻訳	品詞
1	displacement	移動	
2	cavity	空洞	
3	gearmotor	歯車モータ	
4	hole	穴	
5	locking element	ロックングエレメント	
6	locking position	ロックング位置	
7	locking teeth	ロックング歯	
8	pulley	滑車	
9	releasing position	解放位置	
10	rotating pin	回転ピン	

[文献#12 : EP2140068]

	重要技術用語	日本語人手翻訳	品詞
1	perimeter security barrier	周辺防御壁	
2	end post	端柱	
3	intermediate post	間柱	
4	energy absorption means	エネルギー吸収手段	
5	shear base mechanism	剪断基部機構	
6	high tensile steel	高張力鋼	
7	goods vehicle	貨物用車両	
8	impacting vehicle	衝突車両	
9	heavy vehicle	大型車両	
10	wire rope	ワイヤロープ	

[文献#13 : US2008000540]

	重要技術用語	日本語人手翻訳	品詞
1	flexible hose	可撓性ホース	
2	charge air hose	給気ホース	
3	pressure support	圧力支持体	
4	pressure support element	圧力支持部品	
5	anti-adhesion coating	接着防止コーティング	
6	mobility	可動性	
7	strength	強度	
8	bond	接合	
9	elastomer layer	弾性層	
10	matrix	母材	

[文献#14 : EP1738119]

	重要技術用語	日本語人手翻訳	品詞
1	cooling device	冷却装置	
2	control method	制御方法	
3	compressor	圧縮器	
4	condenser	凝縮器	
5	evaporator	蒸発器	
6	suction part	吸入部	
7	pumping part	圧縮部	
8	refrigerant	冷媒	
9	solenoid valve	電磁弁	
10	by-pass line	バイパスライン	

[文献#15 : US2009000389]

	重要技術用語	日本語人手翻訳	品詞
1	air mass flow	空気流量	
2	sensor assembly	センサアセンブリ	
3	internal combustion engine	内燃機関	
4	resistance element	抵抗素子	
5	air intake duct	吸気ダクト	
6	multiple path	多重流路	
7	inlet duct	吸入ダクト	
8	sensing duct	検出ダクト	
9	bypass duct	バイパスダクト	
10	damper	ダンパー	

[文献#16 : US2009000166]

	重要技術用語	日本語人手翻訳	品詞
1	safety sign	安全標識	
2	stop sign	停止標識	
3	lighted sign	点灯式標識	
4	internal lighting	内部照明	
5	front panel	前面パネル	
6	reflective panel	反射パネル	
7	cut out letter	切り抜かれた文字	
8	ambient light	周囲光	
9	light source	光源	
10	solar panel	ソーラーパネル	

[文献#17 : EP1872211]

	重要技術用語	日本語人手翻訳	品詞
1	container-managed application	コンテナ被管理アプリケーション	
2	application program interface	アプリケーション・プログラム・インターフェース	
3	service container	サービス・コンテナ	
4	pluggable interface	プラグ可能なインターフェース	
5	shared service	共有サービス	
6	common service	共通サービス	
7	application process	アプリケーション・プロセス	
8	service discovery API	サービス探索API	
9	processing device	処理装置	
10	search	検索	

[文献#18 : US2010000992]

	重要技術用語	日本語人手翻訳	品詞
1	microwave steamer	マイクロ波蒸し器	
2	cooker container	調理容器	
3	boiler container	ボイラ容器	
4	slip fit	スリップフィット	
5	arcing	アーク放電	
6	offset	オフセットした	形容詞
7	grill shaped	焼き網状の	形容詞
8	snug slip fit mating engagement	ぴったりとしたスリップフィット係合	
9	machine washable	洗濯機洗い可能な	形容詞
10	air hydraulic press	空気液圧プレス	

[文献#19 : EP2269220]

	重要技術用語	日本語人手翻訳	品詞
1	solder bump	はんだバンプ	
2	functional domain	機能領域	
3	footprint packaging	フットプリントパッケージング	
4	self-contained systems-on-chip	自己内蔵型システムオンチップ	
5	radio-base band SoC	無線ベースバンドS o C	
6	physical rearrangement	物理的再配置	
7	interconnect routing	相互接続配線	
8	ESD crowbar circuit	E S D クローバ回路	
9	redistribution layer	再配線層	
10	in situ	原位置で	副詞

[文献#20 : EP2401878]

	重要技術用語	日本語人手翻訳	品詞
1	hand off process	ハンドオフ処理	
2	base station	基地局	
3	subnet prefix	サブネット・プリフィックス	
4	care-of-address	気付けアドレス	
5	ping packet	ピング・パケット	
6	connectivity	接続性	
7	mobility	移動	
8	suffix	サフィックス	
9	handover process	ハンドオーバー処理	
10	node	ノード	

英日テストセット データサンプル

[illegible]

添付資料6-1

[illegible]

特許庁機械翻訳評価手法 評価作業手順書

「特許庁機械翻訳評価手法」による評価作業は、機械翻訳結果を対象とし、以下の4種類の評価を行います。

- (1) 流暢さの評価
- (2) 内容の伝達レベルの評価
- (3) 重要技術用語の翻訳精度の評価
- (4) チェックリストに基づく評価

機械翻訳結果は、1つの文に対しn種類ずつ存在します。1文あたりの全ての機械翻訳結果を、同一の評価者が担当してください。また、1文あたりの全ての機械翻訳結果を連続して評価し、相対的な評価を行ってください。

機械翻訳結果される原文はさまざまな技術分野に属する特許文献から採取されています。評価は、それぞれの技術分野に十分な知識を有する評価者が行ってください。

以下、各評価の基準及び作業手順を示します。

1. 作業用ファイル

評価対象である機械翻訳結果は、その原文及び人手翻訳とともに「作業用ファイル」として示されます。

評価作業は、本ファイル上で行い、評価結果をファイルに入力してください。

次ページに、作業用ファイルの構成を示します。

※作業用ファイルに関する注意

- ・ 作業環境に応じ、列幅や列の表示／非表示を変更可能です。
- ・ 評価結果入力欄およびメモ欄以外は、セルの内容の変更はできません。
- ・ 原文欄をダブルクリックすると、原文に含まれる主な技術用語の辞書引き結果が参照できます。ただし、辞書引き結果はあくまで参考情報であり、提示された訳語が必ずしもその文の技術内容に照らして適訳であるとは限りませんのでご注意ください。
- ・ 「(4) チェックリストに基づく評価」の各カウント項目欄は、該当する翻訳ミスがない場合は空欄のままで結構です。
- ・ 「追加カウント項目」欄及び「英日カウント項目」欄では、対象文が該当する可能性のない項目がグレイアウトされています。ただし、グレイアウトされている項目であっても、カウントが必要と判断した場合はカウントしてください。セルに着色しているだけです、入力が可能です。
- ・ 「機械翻訳結果」欄には、原文に対応する機械翻訳結果を1セル中に収める都合上、

改行で区切られた文は連結しています。ただし、改行が妥当な位置（例えば段落番号や「:」直後）でなく、文の途中での強制改行等によるものの場合は、改行箇所を「///」で示しています。

- ・ 「原文」欄、「人手翻訳」欄、および「機械翻訳」欄では、重要技術用語として指定された各語、およびその人手翻訳が赤フォントでハイライト表示されています。ただし、名詞以外の語や、名詞であっても特殊な複数形で出現している場合等、ハイライト表示されない場合もありますのでご注意ください。

※作業用ファイル構成

項目名	セル	備考
No./文番号	A/B	シリアル番号/原文のシリアル番号
文献番号	C	原文を採取した特許文献の番号
対応番号	D	人手翻訳を採取した対応特許文献の番号
項目	E	原文の記載箇所を Title、Abst、Claim、Desc で表示
IPC	F	原文の筆頭 IPC のクラス
原文	G	機械翻訳対象となった原文
語数/文字数	H	原文の語数
人手翻訳	I	原文に対応する人手翻訳文。対応特許文献から採取し、必要に応じて修正している。
機械翻訳 ID	J	機械翻訳エンジンを区別するための ID
機械翻訳結果	K	評価対象となる機械翻訳結果
メモ欄	L	☆自由使用欄
流暢さ	M	★「(1) 流暢さの評価」の評点を、5～1（半角数字）でセットしてください。
内容伝達レベル	O	★「(2) 内容の伝達レベルの評価」の評点を、5～1（半角数字）でセットしてください。
重要技術用語①～③	O～P、R～S、 U～V	「(3) 重要技術用語の翻訳精度の評価」の対象となる技術用語と、これに対応する人手翻訳を、最大 3 語までセット
用語①～③評価	Q、T、W	★「重要技術用語①」～「同③」それぞれに対する評価値を A～D（半角英字）でセットしてください。
用語数	(X)	（重要技術用語として選定された技術用語数）※非表示
必須カウント項目	Y～AE	★「(4) チェックリストに基づく評価」の【4-1 言語共通項目】中、【必須カウント項目】A～G に該当する翻訳ミスのカウントし、半角数字で入力してください。
A～G 合計	AF	必須カウント項目 A～G の合計カウント数を自動計算
追加カウント項目	AG～AR	★同【追加カウント項目】Z-00～Z-11 に該当する翻訳ミス数を、半角数字で入力してください。
英日項目	AS～BH	★同【英日項目】EJ-01～EJ-16 に該当する翻訳ミス数を、半角数字で入力してください。

2. 各評価基準の説明と作業手順

(1) 流暢さの評価

まずは機械評価結果のみを読んで、日本語として自然な文であるかを以下の 5 段階の評価基準で評価してください。

- 5：文意が明解で、人間が書いた日本語文に近い。
- 4：日本語文として不自然な箇所を若干含むが、文意は明解である。
- 3：日本語文として不自然な箇所があり、文意がわかりにくい。
- 2：日本語文法規則に反する表現をかなり含む。文意がわからない。
- 1：日本語文として成立していない。

流暢さの評価は、機械翻訳結果が日本語として自然な文・表現であるかのみを評価します。従って、上記基準における「文意」とは、あくまでも機械翻訳結果において読み取れる文意を指し、原文や人手翻訳の内容と一致しているかどうかは考慮しません。

※添付資料 1 に、「流暢さの評価」の評価サンプルを示しましたので併せてご確認ください。

(2) 内容の伝達レベルの評価

機械翻訳結果が、原文の内容をどの程度正確に伝達しているかを、人手翻訳の内容に照らして以下の 5 段階の評価基準で主観的に評価してください。

- 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%)

ここでいう「重要情報」とは、原文に含まれる全ての実質情報（技術的要素とその相互関係）を指します。重要情報それぞれの重要度と翻訳の精度から、総合的な評価を行ってください。各評点に付記したパーセンテージは、重要情報が正確に伝達される度合いの大体の目安です。

評価結果を、作業ファイルの「内容伝達レベル」欄に半角数字（5～1）でセットしてください。

なお、原文において複数の解釈が可能な表現であっても、人手翻訳の内容の解釈と異なるものは不正確とみなしてください。

※添付資料 2 に、「内容の伝達レベルの評価」の評価サンプルを示しましたので併せてご確認ください。

(3) 重要技術用語の翻訳精度の評価

原文に含まれる重要技術用語が、機械翻訳結果において適切に翻訳されているかを評価します。指定された重要技術用語それぞれに対し、技術分野 (IPC) 及び人手翻訳に照らし、以下の評価基準 A～D で評価してください。

- 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はないが、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。

機械翻訳結果 1 文において指定される重要技術用語は 0 語～3 語であり、0 語の場合を除き、作業ファイルの「重要技術用語①～③」欄に、原語と人手翻訳がそれぞれ示されています。評価結果を、対応する「用語①～③評価」欄に半角英字 A～D で入力してください。

※添付資料 3 に、「重要技術用語の翻訳精度の評価」の評価サンプルを示しましたので併せてご確認ください。

(4) チェックリストに基づく評価

機械翻訳結果を原文・人手翻訳と比較し、全ての翻訳ミスについて、チェックリストに基づきカウントしてください。チェックリストは、言語共通項目である【必須項目】と【任意項目】、そして【翻訳言語別任意項目】にて構成されます。以下に、それぞれのチェックリスト項目を示します。

【4-1 言語共通項目】

言語によらず機械翻訳全般にて発生する翻訳ミスをチェックするための項目です。【必須項目】と【任意項目】に区分されます。【任意項目】は、評価の目的等に応じて必要な項目のみが評価対象として選定されます。

【必須項目】

- A. 語の欠落
- B. 不要語の混入
- C. 未知語
- D. 品詞の誤り
- E. 係り受け誤り
- F. 語の誤訳
- G. その他の翻訳ミス

【任意項目】

※上記 A～G でカウントした翻訳ミスについて、特に下記にあてはまるものはこちらにもカウントする。

- Z-00・訳語のゆれ
- Z-01・不適切な文の分断
- Z-02・主語・述語・目的語（他動詞の場合）の相互関係に関する翻訳ミス
- Z-03・並列表現に関する翻訳ミス
- Z-04・発明の名称など、名詞句を文として扱った翻訳ミス
- Z-05・特許特有の用語や表現に関する翻訳ミス
- Z-06・特許特有の記号や文献番号表記、付番に関する翻訳ミス
- Z-07・箇条書き表現に関する翻訳ミス
- Z-08・原文のカッコ書きに起因する翻訳ミス
- Z-09・数式、化学式に関する翻訳ミス
- Z-10・数値単位に関する翻訳ミス
- Z-11・数値範囲に関する翻訳ミス

機械翻訳結果に含まれる全ての翻訳ミスをチェックリストと照合し、まずは作業ファイルの【必須項目】欄 A～G に、各チェック項目に該当する翻訳ミス数を半角数字で入力してください。さらに、翻訳ミスが【任意項目】の各チェック項目に該当する場合は、同欄 Z-00～Z-11 にもカウントしてください。

翻訳ミスは、語単位（複合語やイディオムは 1 語とみなします）、係り受け関係単位でカウントします。「(2) 重要技術用語の翻訳精度の評価」で評価対象となった語も再度評価対象としてください。

1 つの翻訳ミスは、チェック項目 A～G の 1 項目のみにカウントします。複数の項目に該当する場合は、上位項目（A>G）のみにカウントしてください。ただし、1 語に対して複数の翻訳ミスが生じている場合は、個別にカウントします。

【任意項目】は、チェック項目 A～G との重複カウント項目となります。任意項目（Z-00

～Z-11) 間の重複カウントも可とします。

なお、【任意項目】に関しては、評価対象となる機械翻訳結果において該当する可能性がないと考えられるチェック項目は、作業ファイルのセルがグレイアウトされています。ただし、グレイアウトされた項目であっても、カウントが必要と判断した場合はカウントしてください。

日本語として読みにくい表現であっても、意味的・文法的に誤解を与える懸念の少ない語や語順、文構造は、「(1) 流暢さの評価」で評価し、ここでは翻訳ミスと見なしません。

※添付資料 3 に、「チェックリストに基づく評価【言語共通項目】」の詳細説明を示しましたので併せてご確認ください。

【4-2 翻訳言語別共通項目】

特定の翻訳言語に特有の主要な翻訳ミスをチェックするための項目です。評価の目的等に応じて必要な項目のみが評価対象として選定されます。

以下に、英日機械翻訳に対する翻訳言語別共通項目（「英日項目」）の内容を示します。

- EJ-01・文中のピリオドを適切に処理していない
- EJ-02・「:」や「;」を用いた文を適切に処理していない
- EJ-03・重要な単数・複数情報が維持されていない
- EJ-04・単数形・複数形併記表現を適切に処理していない
- EJ-05・序数詞表現が正しく解釈されていない
- EJ-06・慣用句の中間への語の挿入を適切に処理していない
- EJ-07・倒置文を適切に翻訳していない
- EJ-08・名詞と代名詞の関係を適切に表現できていない
- EJ-09・慣用表現とみなすべき場合／そうでない場合を適切に翻訳できていない
- EJ-10・前置詞に対し適切な訳語を選択できていない
- EJ-11・接頭辞、接尾辞を伴う語を適切に処理していない
- EJ-12・スラッシュを用いた並列表現を適切に処理していない
- EJ-13・therein, therefor, wherein, 等を適切に処理していない
- EJ-14・代名詞 same、the same を適切に処理していない（【発明の名称】）
- EJ-15・小文字を含む大文字表記を適切に処理していない（【発明の名称】）
- EJ-16・クレーム引用に関する表現を適切に翻訳していない（【特許請求の範囲】）

言語共通項目における【任意項目】と同様、翻訳ミスが上記の各チェック項目に該当する場合に限り、作業ファイル「英日項目」欄の EJ01～EJ16 にカウントしてください。

「英日項目」EJ01～EJ16 間での重複カウント、及び【任意項目】との重複カウントも

可とします。

なお、評価対象となる機械翻訳結果において該当する可能性がないと考えられるチェック項目は、作業ファイルのセルがグレイアウトされています。ただし、グレイアウトされた項目であっても、カウントが必要と判断した場合はカウントしてください。

※添付資料 4 に、「チェックリストに基づく評価【英日項目】」の詳細説明を示しましたので併せてご確認ください。

以 上

添付資料 1 : 「流暢さの評価」の評価サンプル

「流暢さの評価」では、機械翻訳結果を単独の文として読んだ際の読みやすさ、文としての表現の自然さのみを主観的に評価する指標です。原文の内容を正しく反映しているかどうかは不問であるため、原文・人手翻訳に照らす必要はありません。

流暢さの評価基準は、以下の 5 段階です。

- 5 : 文意が明解で、人間が書いた文に近い。
 4 : 文として不自然な箇所を若干含むが、文意は明解である。
 3 : 文として不自然な箇所があり、文意がわかりにくい。
 2 : 文法規則に反する表現をかなり含む。文意がわからない。
 1 : 文として成立していない。
 ※なお、ここでの評価は、翻訳結果のみを単独の文として扱う。「文意」とは、あくまで翻訳結果において読み取れる文意を指し、原文やリファレンス文の内容と一致しているかどうかは考慮しない。

以下に、各評点の程度を表すサンプルを示します。あくまで、文としての自然さのみが評価観点となります。

5	文意が明解で、人間が書いた文に近い。
	好ましくは、油圧シリンダーが上部のアーム組立品を操作するために使用されている一方で、電動アクチュエータは脚拡張および側面のエクステンション・アームを操作するために使用される。
4	文として不自然な箇所を若干含むが、文意は明解である。
	装置は、3 点のカップリング構成で異なる絶頂および幅を持っている様々な道具の用途に適していて、カテゴリ 1 およびカテゴリ 2 道具の用途に特に適している。
3	文として不自然な箇所があり、文意がわかりにくい。
	カップリング組立品 40 は、上部のアーム油圧シリンダー 13 の操作により道具に比べて適切に角度がある。
2	文法規則に反する表現をかなり含む。文意がわからない。
	垂直のものがスリーブをつけることを特徴とする実施の形態において、51 および 61、ならびに拡張脚 52 と 53 角度、内部へ、この最終操作は、使用の間に、よりよく道具 91 を保持する安全な方法で、道具 91 のピン 93 および 94 を固定する役目をする。
1	文として成立していない。
	図。 3、クロススリーブ部材と横延長アーム、及び垂直スリーブ部材、垂直延長脚と下側フック部材を含むほぼ垂直アセンブリを含むほぼ水平立体を示す結合アセンブリの背面図であり、トラクタ下部ヒッチに接続されて示さアーム。

◎補足説明

【内容のナンセンスさの反映】

「流暢さの評価」では、内容の技術的なナンセンスさは減点対象となりません。あくまで、文法的に不自然であるか否かを評価します。

機械翻訳	[0009] 図 3 は、図 1 のマイクロ波汽船の 眼瞼 の平面視である。
評 点	4
評価根拠	「汽船」と「眼瞼」の組み合わせは技術的にはナンセンスだが、ここではそれは問わない。「平面視」という訳語は日本語として不自然であるため、そのみが減点対象。

【機械翻訳がカッコ書きを用いる場合】

日本語文として自然な範囲を逸脱するようなカッコ書きは減点対象としてください。

機械翻訳	2. 請求項 1（第 2 のノズルアレーのノズルが第 1 のノズルアレーのピッチ P に等しいピッチで第 2 のノズルアレーに沿って間隔を置かれる）に記載の流体はじき出しデバイス。
評 点	4
評価根拠	文法的には破綻はないが、カッコ書きが全体の半分以上を占める構成は日本語文として不自然であり、－1 とした。実際は、「請求項 1」とカッコ書きの内容は対応しないため誤訳であるが、流暢さの評価においてはそこを考慮する必要はない。

【未知語を含む場合】

未知語は「日本語文法規則に反する表現」に含まれますので、減点対象としてください。

機械翻訳	しかしながら、上記の通りに、特許 Literature 3 において開示される分離機能層は、単層からなるすべてのポリアミド膜である。
評 点	3
評価根拠	未知語 Literature が含まれていなければ、「上記の通りに」の「に」や、「すべて」の配置程度の軽微な問題しかないので評点 4 が妥当だが、未知語を含むため－1 とした。

以 上

添付資料 2：「内容の伝達レベルの評価」の評価サンプル

「内容の伝達レベル」の評価では、機械翻訳結果が原文の実質的な内容をどの程度正確に伝達しているかを、人手翻訳の内容に照らして主観的に評価し、下記の 5 段階評価で評価します。

- 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%)

「重要情報」とは、原文に含まれる全ての実質情報（技術的要素とその相互関係）を指します。評価は、文に含まれる全ての重要情報を対象に、それぞれの重要度と翻訳精度を考慮して、機械翻訳結果の内容の伝達レベルを総合的・主観的に行なってください。なお、各評点に付記したパーセンテージは、機械翻訳結果が原文の意味をどの程度正確に伝達しているかを示す大体の目安であり、「重要情報」の個数から厳密に算出する必要はありません。

1. 評価サンプルと評価根拠

以下、サンプル文を用いて、「重要情報」の捉え方と、各基準に相当する機械翻訳結果の実例ならびにその評価根拠を示します。

【サンプル文】

原 文	The distribution space 33 is sealed with a gasket 26 along the perimeter.
人手翻訳	分配スペース 3 3 は、その外周に沿ってガスケット 2 6 により密閉される。

上記サンプル文における重要情報は、技術内容に関する各技術要素（分配スペース 3 3、外周に沿って、ガスケット 2 6、密閉される）と、それぞれの要素の相互関係となります。このうちで最も重要度の高い情報は「分配スペースがガスケットにより密閉される」という骨格構造に含まれる各技術要素及びその相互関係といえます。一方、「分配スペースの外周に沿って（密閉される）」という情報は、これも実質情報であり評価対象ですが、そこに含まれる技術要素及び各相互関係の重要度は前記各要素よりは低くなります。評価にあたっては、こうした重要情報それぞれの重要度の違いや訳語の精度も考慮して総合的に評価値を決定してください。

なお、本項目において「実質的でない」情報とは、例えば能動態・受動態の言い換えや語順等、内容の伝達に影響しない情報を指し、前者については例えば「ガスケットが分配スペースを密閉する」という言い換えや、後者については「ガスケットにより、分配スペース 3 3 は密閉される」という順序の違いは不問としてください。

以下に、上記サンプル文について、評点 5～1 それぞれに該当する機械翻訳結果のサンプルを示します。なお、「評価理由」欄に示したパーセンテージは、機械翻訳結果の内容から総合的に判断した主観的な数値です。

【評点別の評価サンプル】

点	機械翻訳結果	評価理由
5	分配空間 3 3 は、外周に沿ってガスケット 2 6 により封止されている。	全ての重要情報が正しく把握でき、減点すべき箇所なし。伝達レベルは 5（ほぼ 100%）と評価した。
4	分配空間 33 は周囲とともにガスケット 26 で密閉される。	「外周に沿って」という情報が把握できないが、各要素の訳語と文の骨格構造は正しい。大枠は理解できるので、伝達レベルは 4（80%強）とした。
3	分配スペース 33 は視野計に沿ってガスケット 26 で封をされる。	文の骨格構造や大半の要素の訳語は正しい点は評点 4 の例と同じだが、「外周」→「視野計」と大きく誤訳されており、正しい意味の類推が困難かつ不要な情報が与えられているため、より低い評価とし、伝達レベルは 3（70%程度）とみなした。
2	流通空間 3 3 は、周辺部に沿ったガスケット 2 6 を有する溶封である。	文の主語が「流通空間」と意味不明の訳語となっているほか、述語「溶封」も不正確で誤解を招く。文の骨格構造もわかりにくく、伝達レベルは 2（30%程度）と評価した。
1	分布に沿ってガスケット 2633 周囲密閉空間	「分配スペース 33」が「分布」「空間」「33」と分断されており把握不可能。各要素の相互関係も理解不可能であり、正しく把握できる情報は「ガスケット」のみ。伝達レベルは 1（5～10%程度）と判断した。

2. 文の骨格構造（主語・述語・目的語）の誤り

前述のとおり、本評価の対象は、文に含まれる全ての技術的要素とその相互関係であり、これらが機械翻訳結果に正しく伝達されているかを、それぞれの重要度ならびに訳語の精度から総合的に判断し、評点を決定します。

従って、文の骨格構造が大きく誤っている場合でも、無条件で最低評価（評点 1）となるわけではなく、個々の構成要素等に正確に訳されているものがあれば、それに応じて評点 2 以上が付与される場合もあります。

ただし、文の骨格構造が大きく誤訳された結果、文脈や構成要素等を踏まえても本来の文の主旨を類推不可能であるような場合は、評点は 2 以下としてください。また、読み手に文の主旨を誤解させる可能性の高い誤りは、程度によっては、「(本来の) 文意がわから

ない」ものとし、仮に構成要素の 20%以上が適切に訳されていたとしても、評点 1 が妥当な場合もあります。例えば、以下のような場合です。

原 文	Provided is an emission treatment system and method for simultaneously remediating the nitrogen oxides (NO _x), particulate matter, and gaseous hydrocarbons present in diesel engine exhaust streams.
人手翻訳	ディーゼルエンジン排気流れに存在する窒素酸化物（NO _x ）と粒状物と気体状炭化水素を同時に修正するに適した 排気処理システムおよび方法を提供する 。
機械翻訳	排気処理システムおよび方法は、同時に窒素酸化物（NO _x ）を是正するため、粒子状物質、及びディーゼルエンジン排気流中に存在する ガス状炭化水素を提供する 。
評 点	1
評価根拠	技術用語は概ね適切に訳されているものの、文の骨格構造は完全に誤っており、発明の主題が「廃棄処理システムおよび方法」ではなく「ガス状炭化水素」とであると誤解される訳文となっている。このような文は、「骨格構造が大きく誤って」おり、「本来の文の主旨を類推」することが困難、かつ「読み手に文の主旨を誤解させる可能性」が非常に高いため、「(本来の) 文意がわからない」ものとし、評点 1 とした。

3. 誤解を招く誤訳

誤訳自体が本質的に誤解を招くものであることは確かですが、一見して明らかに誤りと分かるような誤訳ではなく、読み手がそれと気づきにくい誤訳は、より大きな悪影響を実用の際に及ぼす危険性があります。このような「誤解を招きやすい」誤訳は、普通の誤訳よりも訳語の精度が悪いものとして扱ってください。

原 文	Non-ferromagnetic layer 104 may include, for example a conductive material layer having a thickness in a range of 1-3 nm, although non-ferromagnetic layer 104 may be thicker or thinner, in other embodiments.
人手翻訳	非強磁性層 104 は、たとえば厚さが 1～3 nm の範囲の導電性材料層を含み得るが、他の実施例では非強磁性層 104 はより厚い場合もあれば、より薄い場合もある。
機械翻訳	非強磁性層 104 は、他の実施形態では、より厚い又はより薄いかもしれないが、非強磁性層 104 は、例えば、 1～20nm の範囲の厚さを有する導電性材料層を含むことができる。
評 点	3
評価根拠	数値範囲「1～3nm」が「1～20nm」と誤訳されている。原文と照らさない限り、読み手がこの誤訳に気づく可能性は低く、従って大きな誤解を与える危険性が高い。文の骨格構造は（流暢ではないが）正確であり、また各構成要素の訳語も適切で、誤訳は上記数値範囲部分だけであるが、重要な構成要素であり、また訳語の精度が非常に悪いため、文意の少なくとも 20%以上を損ねていると判断し、評点は 3 が妥当と判断した。

4. 特許知識・経験により正確に類推できる誤訳

評価者ご自身の特許知識・経験により正確に類推できる誤訳については、補正した内容にて評価していただいて結構です。

原 文	13. The method of claim 1 , wherein the powder coating composition comprises a polymer selected from the group consisting of an acrylic, polyester, polyurethane, polyether, polyvinyl, cellulosic, acrylate, silicon-based polymers, co-polymers thereof, and combinations thereof.
人手翻訳	【請求項 1 3】 前記粉末被覆組成物が、アクリル、ポリエステル、ポリウレタン、ポリエーテル、ポリビニル、セルロース、アクリレート、ケイ素系ポリマー、これらのコポリマー及びこれらの組合せからなる群から選択されるポリマーを含有する、 請求項 1 記載の方法。
機械翻訳	13. 請求項 1 (粉末コーティング組成物がアクリル、ポリエステル、ポリウレタン、ポリエーテル、ポリビニル、セルロース誘導体、アクリレート系繊維、シリコン系ポリマー、そのコポリマーおよびそれらの組み合わせから成る群から選択されるポリマー) の方法。
評 点	4
評価根拠	原文は、請求項 1 を引用した引用請求項だが、機械翻訳結果は、引用請求項の内容を請求項 1 の内容として扱っており、大きな誤解を生む。ただし特許文献に通じている者にとっては、引用請求項にて被引用請求項の内容のみが示されることはありえないことは自明であり、これを正しく補正して解釈することが可能である。このため、機械翻訳結果の内容は不正確だが、それを補正して得られる情報は正確である。係り受けに誤りがあることは確かだが、技術内容部分ではないため重要度は低く、かつ知識による補正が可能なものであるため、その影響度はごく軽微。他に大きな誤りはないため、評点 4 (80% 強) が妥当である。

5. 評価の実例集

【サンプル①】

原 文	After the wash has dried, the motor is again powered in the reverse direction to lower the frame for unloading.
人手翻訳	洗濯物が乾いた後に、外す作業のためにモーターを再び逆の方向に作動させてフレームを下ろす。
機械翻訳	洗濯が乾燥した後に、モータは、荷卸でフレームを下ろすために再び逆方向で動かされる。
評 点	4
評価根拠	「モータを逆の方向に作動させてフレームを下ろす」という骨格構造は「モータはフレームを下ろすために逆方向で動かされる」と言い換えられているが意味は同一であり問

	題ない。「(洗濯物を) 外す作業」が「荷卸」と誤訳されているが、類似の行為であり、他の要素が適切に訳されているため正しい意味の類推は可能。他にも「洗濯物」が「洗濯」、「(外す作業) のために」が「で」と曖昧な表現になっているが、誤訳の程度、重要度ともにさほど高くないため大きなマイナスとはならず、伝達レベルは 90%程度と判断した。
--	--

【サンプル②】

原 文	[0006] Usually this pulley is fixed to the rotating pin either with permanent fixing means (screws, washers, etc.) or by means of a tongue arranged on the rotating pin.
人手翻訳	【0006】普通、この滑車は、永久固定手段（ねじ、座金など）で、あるいは回転ピンに配置されたトングによって回転ピンに固定される。
機械翻訳	0006 通常、このプーリに常時固定手段（ねじ、ワッシャなど）と、回転ピンに配置された舌の手段によって回転ピンに固定されている。
評 点	3
評価根拠	骨格構造は「滑車は、永久固定手段あるいはトングによって回転ピンに固定される」となる。「滑車」が「プーリ」と別の訳語で訳されているが同義・同等であり問題なし。一方、「舌の手段」は「トング」の同義語とはいえ減点対象。要素間の相互関係に関しては、「滑車は…固定される」が「プーリに…固定される」と誤って訳されており、重大。また「常時固定手段」と「舌の手段」が or ではなく and で結ばれており、これも重大である。以上に鑑み、全体的な伝達レベルは 50%程度と判断した。なおパラグラフ番号の処理が不完全だが、文意には影響しない形であるため不問とした。

【サンプル③】

原 文	The microwave steamer 10 includes a boiler container 12, a cooker container 14 and a lid 16.
人手翻訳	マイクロ波蒸し器 10 は、ボイラ容器 12 と、調理容器 14 と、蓋部 16 とを含む。
機械翻訳	超短波電波汽船 10 はボイラーコンテナ 12、料理道具コンテナ 14、およびリッド 16 を含んでいる。
評 点	2
評価根拠	文の骨格構造は正しいが、主要素「マイクロ波蒸し器」が「超短波電波汽船」と大きく誤訳されており、技術内容を理解することは困難。また、「調理容器 cooker container」は「調理用容器」の意味だが、「料理道具コンテナ」では料理道具を入れる容器と誤解される。結果、正確に把握できる情報は、要素列挙文である点と、「ボイラーコンテナ」「リッド」程度であり、「超短波電波」に関する技術であることが把握できる点を加味しても、伝達レベルは 30%程度。

【サンプル④】

原 文	They are of solid construction, usually of concrete blocks, making them unsightly and difficult to install.
人手翻訳	それらは中実の構成であり、通常はコンクリートブロックであるために、見苦しくかつ設置が困難なものとなっている。
機械翻訳	彼らは、インストールするには、それらを見苦しいと難しく、通常はコンクリートブロックに、固体の構造である。
評 点	1
評価根拠	文の構造、要素の相互関係が完全に崩れており、「コンクリートブロック」や「見苦しい」、など正しい訳語もあるものの、文意の把握はほぼ不可能。伝達レベルは 10%未満と判断した。

【評点 5 のサンプル集】

原 文	The struts 202 are generally positioned in close proximity upstream of the fan 104, near the outer rim 105 of the air duct 102.
人手翻訳	支柱 202 は、概して、空気ダクト 102 の外側リム 105 の近くでファン 104 の上流に近接して配置される。
機械翻訳	支柱 202 は、一般に、空気ダクト 102 の外縁 105 の近くに、ファン 104 の上流側に近接して配置されている。

原 文	Then, the core/shell growth solution is heated to a temperature sufficient to produce the PbS shell.
人手翻訳	次に、コア／シェル成長溶液を PbS シェルを生成させるために十分な温度に加熱する。
機械翻訳	そして、コア／シェル成長溶液は、PbS シェルを生成するのに十分な温度に加熱される。

原 文	46. The method of any of claim 43-45, wherein the sequence patterns are nucleic acid motifs, amino acid motifs or protein domains.
人手翻訳	【請求項 46】 配列パターンが核酸モチーフ、アミノ酸モチーフ又はタンパク質ドメインである、請求項 43-45 のいずれかに記載の方法。
機械翻訳	46。配列パターンは、核酸モチーフ、アミノ酸モチーフまたはタンパク質ドメインである、請求項 43-45 のいずれかに記載の方法。

【評点 4 のサンプル集】

原 文	A) providing an ink jet printer that is responsive to digital data signals;
人手翻訳	A) デジタルデータ信号に応答するインクジェットプリンターを用意すること
機械翻訳	デジタルデータ信号に反応するインクジェット印字装置を提供する A);

原 文	The difference is then sent to the level controller 128 to adjust the predetermined batch fill rate 246.
人手翻訳	次に、この差はレベルコントローラ 128 に送られ、予定バッチ充填速度 246 が調整される。
機械翻訳	差は、所定のバッチ充填速度 246 を調整するレベルコントローラ 128 に送られる。

原 文	Curve 706 represents the variation in NOx formation with increasing ratio of diluent to fuel.
人手翻訳	曲線 706 は、希釈剤 - 燃料比の増大に伴う NOx 生成の変化を表す。
機械翻訳	曲線 706 は、希釈剤と燃料の比の増加と共に、NOx 生成における変化を表す。

【評点 3 のサンプル集】

原 文	The collection member or members 55 may also be identified as a manifold.
人手翻訳	収集部材 55 は多岐管としても認識され得るものである。
機械翻訳	収集部材または部材 55 は、マニホールドと確認されることもできる。

原 文	Smaller movements of line 224 are ignored or are interpreted in the normal manner for guidance control.
人手翻訳	それよりも小さなライン 224 の動きは、無視するか、またはガイダンス制御について通常の態様で解釈する。
機械翻訳	224 行目のより小さな動作は無視されるか、あるいは誘導制御用の正常な方法で解釈される。

原 文	One end of the suspension 203 is secured to the arm 204, for example, with a ball connecting structure.
人手翻訳	サスペンション 203 は、一端がアーム 204 に、例えばボール接続構造によって固定されている。
機械翻訳	サスペンション 203 の一端は、アーム 204 に固定され、例えば、ボール連結構造を有する。

【評点 2 のサンプル集】

原 文	Combustor 14 channels combustion gases to turbine 16 wherein gas stream thermal energy is converted to mechanical rotational energy.
人手翻訳	燃焼器 14 は、燃焼ガスをタービン 16 へと導き、そこでガス流の熱エネルギーが機械回転エネルギーに変換される。
機械翻訳	燃焼タービン 16 ~14 チャンネルの燃焼ガスのガス流の熱エネルギーが機械的回転エネルギーに変換される。

原 文	These hydroxyl radicals will attack, oxidize, and destroy the toxic agents present on the surface.
人手翻訳	これら水酸基ラジカルは、表面に存在する毒性の物質を攻撃し、酸化させ、且つその毒性を破壊する。
機械翻訳	これらのヒドロキシルラジカルは、攻撃を酸化し、表面に存在する毒物を破壊します。

原 文	18. An illumination apparatus which illuminates a first surface with light supplied from a light source, the illumination apparatus comprising: the spatial light modulation unit as set forth in claim 1.
人手翻訳	【請求項 18】 光源から供給される光によって第 1 の面を照明する照明装置において、請求項 1 に記載の空間光変調ユニットを備えることを特徴とする照明装置。
機械翻訳	18. 発光源、イルミネーション装置を含むことから供給された光を備えた最初の表面を照らすイルミネーション装置: 請求項 1 で述べられるような空間光変調ユニット。

【評点 1 のサンプル集】

原 文	A third gain block 38 applies a scaling factor G_n to the digitized navigation audio signal.
人手翻訳	第 3 のゲインブロック 38 は、倍率 G_n をデジタル化されたナビゲーションオーディオ信号に適用する。
機械翻訳	3 分の 1 にデジタル化しているナビゲーションオーディオ信号へのスケーリングファクタ系球体腎炎を阻害 38 適用に獲得しなさい。

原 文	The nipper frame 17 is movable back and forth by the reciprocal movement of the nipper shaft 20 so that the front end of the bottom nipper 23 is moved close to and away from the detaching rollers 16.
人手翻訳	ニッパフレーム 17 は、ニッパシャフト 20 の往復回転によって、ボトムニッパ 23 の先端部がデタッチングローラ 16 に対して接近・離間するように前後に揺動するようになっている。
機械翻訳	底少年 23 の前方端部が近接移動であるように、少年フレーム 17 は、少年軸体 20 の相反移動によって前後に移動可能である、に、そしてデタッチング・ローラー 16 から遠方に。

原 文	Method and apparatus for transmitting memory pre-fetch commands on a bus
人手翻訳	バス上のメモリプリフェッチコマンドを送信するための方法および装置
機械翻訳	送信メモリのための方法および装置は、前もってバスに関するコマンドを取って来る。

以 上

添付資料 3：「重要技術用語の翻訳精度の評価」の評価サンプル

「重要技術用語の翻訳精度の評価」では、原文に含まれる重要技術用語が、機械翻訳結果において適切に翻訳されているかを、下記評価基準 A～D で評価します。

- 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はないが、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。

重要技術用語は各文についてあらかじめ指定されており、評価者自身がこれを選定する必要はありません。指定される重要技術用語は 1 文あたり 0 語～3 語です。主に名詞が指定されていますが、名詞以外の用語（動詞、形容詞等）が指定されている場合もあります。

1. 評価サンプル

以下、サンプル文を用いて、重要技術用語の評価の実例とその評価根拠を示します。

【サンプル文】

原 文	In such an embodiment, the coating 12 may comprise a multilayer coating having alternating layers of B4C/B9C.
人手翻訳	かかる一実施形態においては、被膜 12 は、B 4 C / B 9 C の交互層を有する 多層被膜 を含むことができる。
【重要技術用語】：multilayer coating（人手翻訳「多層被膜」）	

[A 評価（適訳語）]

機械翻訳	そのような実施例では、被覆処理な 12 は B4C/B9C の交互層を持っている 多層コーティン グ を含むことができる。
------	--

重要技術用語の訳語が、人手翻訳と同一か、もしくは技術的に同義であり、かつ一般的に用いられる訳語であれば、A 評価（適訳語）とします。

本例の場合、「多層コーティング」以外にも、「多層被覆」や「多層皮膜」が A 評価とすべき訳語として挙げられます。

[B 評価（可訳語）]

機械翻訳	そのような実施例では、覆う 12 は、B4C/B9C の交替する層がある 多層のコーティング を含んでもよい。
------	--

重要技術用語の訳語が、技術用語として一般的に用いられるものではないが、意味はおおむね正しい場合、B 評価（可訳語）とします。

本例の「多層のコーティング」は、意味的には「多層コーティング」と同等ですが、一般的には「多層」と「コーティング」の間に「の」を入れた表現は使われないため、A 評価とはなりません。このように、意味的には同等であっても、一般的な表現でないような表現は、B 評価としてください。B 評価の例としては、「複数層コート」や「多レイヤー皮膜」等も該当します。

【C 評価（誤訳語）】

機械翻訳	このような実施形態では、コーティング 12 は B4C/B9C の交互層を有する 多層を覆うこと を含んでよい。
------	---

重要技術用語が正しい品詞で訳されていないかったり、誤った訳語もしくは技術的に不適切な訳語に訳されている場合は、C 評価（誤訳語）とします。

【D 評価（不訳語）】

機械翻訳	このような実施形態では、コーティング 1 2 は、B4C/B9C の交互層を有する multi layer コーティングを含むことができる。
------	---

重要技術用語またはその一部が未知語扱い（原語のまま出力）であったり、欠落している（翻訳結果中に存在しない）場合は、D 評価（不訳語）としてください。

なお、重要技術用語に相当する訳語が特定できなかった場合も、D 評価として下さい。

【訳語のゆれ】

1 つの重要技術用語が文中に複数回出現し、それぞれの訳語が異なる場合は、**最も良いと思われる訳語の評価値を入力してください。**

【サンプル文】

原 文	[0002] The present disclosure relates generally to a substrate and, more particularly, to a system and method for treating a coating on a substrate .
人手翻訳	[0002] 本開示は、一般に 基板 に関し、より詳細には 基板 上の被膜を処理するシステム及び方法に関する。
機械翻訳	[0002] 本発明は、 基板 に係り、特に、 基体 上の被覆を処理するための方法に関する。

上記例の場合、重要技術用語「substrate：基板」が文中に 2 回出現しており、機械翻訳結果においては 1 回目が「基板」（A 評価）、2 回目が「基体」（B 評価）と翻訳されています。

この場合、評価値は最も良い A とします。

[原語のカタカナ表記]

原語をカタカナ表記しただけの訳語は、厳密には誤訳（C）とはいえませんが、読んで意味が分からないような場合は、それと同等とみなし、C 評価としてください。

読んで意味が分かるものについては、日本語として一般に使われる同義語の範囲であれば適訳（A）、そうでなければ可訳（B）としてください。両者の区別は、概ね「検索の際に同義語として思いつくかどうか」となります。

以下、可訳語（B）と判定するのが妥当な実例をいくつか挙げます。

gas transfer	ガス・トランスファー
melting	メルティング
SHEET FOLDING METHOD	シートフォールディング法
electric field	エレクトリックフィールド

以 上

添付資料 4-1：「チェックリストに基づく評価（言語共通項目）」詳細説明

「チェックリストに基づく評価」は、機械翻訳結果に含まれる全ての翻訳ミスを、チェックリストの項目別にカウントしていく評価です。

1. 翻訳ミスの検出

はじめに、機械翻訳結果を対応する原文・人手翻訳と比較し、そこに含まれる全ての翻訳ミスを検出します。翻訳ミスは、語単位（複合語やイディオムは一語と見なす※）、係り受け関係単位を原則とします。また、重要技術用語として評価した語も再度評価対象となります。

翻訳ミスを語単位、係り受け関係単位に検出する実例を以下に示します。

原 文	Preferably, a contact member or members is mounted to the container spaced from and facing the contact member or members carried by the upright restraint.
人手翻訳	好ましくは、直立固定具に保持された1つ以上の接触部材から離間し、接触部材と対向して1つ以上の接触部材がコンテナに取り付けられる。
機械翻訳	好ましくは、間隔を置いて配置されて、接点部材に向かっているコンテナまたは直立した抑制器具により担持される部材に、接点部材または複数部材は、載置される。

上例における「語単位、係り受け関係単位」の翻訳ミスは、以下の4ヶ所となります。なお、原文中 **contact member or members** は2回出現しますが、翻訳ミスとして指摘されているのは全て2回目のものです。

- ① **mounted to** の目的語は **container** のみだが、**container** と **members** の双方が目的語となっている。
- ② **spaced from** の目的語は **contact member or members** だが、いずれにも係っていない。
- ③ **facing** も **contact member or members** が目的語だが、**contact member** にしか係っていない。
- ④ **carried by** も **contact member or members** を修飾しているが、**members** にしか係っていない。

これら4つの翻訳ミスは、「(2回目の) **contact member or members** が並列と解釈されていない」という1つの翻訳ミスに集約される、という考え方も可能ではあります。しかし、このような解釈では、仮にこれら4つの翻訳ミスのうちの一部が解消したとしても、4つ全てが解消しないかぎり翻訳ミスの個数は1のままとなり、翻訳品質を精密に評価するための指標としては不適當です。このため、上例のとおり、1つの語が複数の係り受け関係を有する場合は、それぞれの係り受け関係ごとに評価を行う必要があります。

なお、翻訳ミスの検出は、係り受け関係にある語のどちら側を基準にしても構いません。上に示した例は係る側を基準にしていますが、係られる側を基準としたとしても、以下のとおり、検出される翻訳ミスの数は同数となります。

- ①members が誤って mounted to の目的語となっている。
- ②contact member or members が spaced from の目的語となっていない。
- ③members が facing の目的語となっていない。
- ④contact member が carried by に修飾されていない。

※翻訳ミスは語単位が原則ですが、複合名詞やイディオムは 1 語扱いとします。複合名詞に該当するか否かの判断は、以下を参考としてください。

・「脂肪アルコールを含む組成物」

- ⇒ fatty alcohol containing composition は 1 語とみなす。
- ⇒ composition containing fatty alcohol は composition、containing、fatty alcohol に分ける。
- ⇒ fatty alcohol contained composition は 1 語とみなす。
- ⇒ fatty alcohol contained in composition は当然ながら fatty alcohol、containing、in、composition に分ける。

2. 必須項目と任意項目

「チェックリストに基づく評価」の「言語共通項目」では、言語を問わず機械翻訳全般において発生し得る主な翻訳ミスを【必須項目】とし、以下のとおり A～G の 7 項目に分類しています。さらに、機械翻訳が特に苦手とする特定のケースや、特許文献に特有または頻出のケースのみを対象とした特許文献の機械翻訳特有の翻訳ミスを【任意項目】として項目 Z に 12 項目を列挙しています。

【必須項目】

- A. 語の欠落
- B. 不要語の混入
- C. 未知語
- D. 品詞の誤り
- E. 係り受け誤り
- F. 語の誤訳
- G. その他の翻訳ミス

【任意項目】

※上記 A～G でカウントした翻訳ミスについて、特に下記にあてはまるものはこちらにもカウントする。

- Z-00・訳語のゆれ
- Z-01・不適切な文の分断
- Z-02・主語・述語・目的語（他動詞の場合）の相互関係に関する翻訳ミス
- Z-03・並列表現に関する翻訳ミス
- Z-04・発明の名称など、名詞句を文として扱った翻訳ミス
- Z-05・特許特有の用語や表現に関する翻訳ミス
- Z-06・特許特有の記号や文献番号表記、付番に関する翻訳ミス
- Z-07・箇条書き表現に関する翻訳ミス
- Z-08・原文のカッコ書きに起因する翻訳ミス
- Z-09・数式、化学式に関する翻訳ミス
- Z-10・数値単位に関する翻訳ミス
- Z-11・数値範囲に関する翻訳ミス

項目 A～G は【必須項目】です。評価者は、検出した各翻訳ミスを、チェック項目 A～G のいずれか 1 ヶ所のみにカウントします。

その後、各翻訳ミスを【任意項目】(Z-00～Z-11) と照合し、該当する場合は、こちらにもカウントします。なお、対象文が該当する可能性がないと考えられる項目のカウント入力欄はグレイアウトされています(※)。

※ グレイアウトされた項目でも、そこに該当する翻訳ミスを検出した場合はカウント値を入力してください。評価値入力欄はグレイアウトされていても入力可能です。

[カウント基準]

項目 A～G へのカウントは、1 つの翻訳ミスについて 1 回のみです。翻訳ミスが複数の項目に該当する場合は、上位項目 (A>G) のみにカウントしてください。

例えば、「D. 品詞の誤り」に該当する翻訳ミスは、常に「F. 語の誤訳」を併発しますし、多くの場合「E. 係り受け誤り」も誘発します（例えば名詞「light : 光」が形容詞「軽い」と品詞誤りされた場合、語の誤訳にも該当し、係り受けも誤ったものとなる可能性が高い）が、このような場合、カウントはこのうち最上位の「D. 品詞の誤り」のみにカウントします。

ただし、1 語に対して複数の翻訳ミスが生じている場合は、それぞれを個別の翻訳ミスとして取扱い、カウントして下さい。以下に、1 語に対して 2 つの翻訳ミスが生じている実例を示します。

原 文	This permits the lower edges of the windscreens to slide over the upper surface of the slide bars 9,10 without hindrance from the brake bars 25,26.
人手翻訳	これによって、フロントガラスの下側の縁がブレーキバー 25、26 によって妨げられずにスライドバー 9、10 の上面を滑ることが可能になる。
機械翻訳	これは、フロントガラスの下方端が制動棒 25、26 からスライド・バー9、10 の上面の無事上を滑るのを許容する。

この例における **without hindrance** は、人手翻訳のとおり「妨げられず」と訳すのが適当ですが、機械翻訳結果では「無事に」となっており、まずはこれを 1 つの翻訳ミス（語の誤訳）としてカウントする必要があります。さらに、この語は「slide : 滑る」に係るべきものであり、語の誤訳があるにせよ少なくとも「無事に滑る」と訳されなければならないところ、機械翻訳結果ではそうっておらず、おそらく「上面 : upper surface」に係っています。従ってこちらも別の翻訳ミス（係り受けの誤り）として再度カウントする必要があります。

先に重複カウント禁止の例として示した事例（「light : 光」）との違いは、今回の事例は、一方の翻訳ミスが他方の翻訳ミスの原因とみなされず、それぞれが独立して発生しているとみなせる点にあります。主に「E. 係り受け誤り」と「F. 語の誤訳」が、それぞれ独立した翻訳ミスとして一語に対し同時に発生することが多くなります。

3. 各チェック項目の説明と具体例

ここでは、必須項目 A～G および任意項目 Z の各チェック項目について、対象となる翻訳ミス等の詳細を具体例にて説明します。

<A. 語の欠落>

原文中の語に対応する訳語が、機械翻訳結果中に存在しないケースを指します。

なお、人手翻訳は原文中にない語が補足されている場合もありますので、必ず原文と機械翻訳結果を対比してください。

原 文	Physiologically, the voltage- difference in the myocardial cells of the heart should be close to 0, because the ventricles were just activated and had contracted (in the QRS).		
人手翻訳	生理学的に、(QRSにおいて) 心室がちょうど活性化され、収縮したので、心筋細胞の 電圧差 は0近くにあるべきである。		
機械翻訳	生理的に、 電圧 —心臓の心筋細胞において、室がちょうど起動して、収縮した (QRS) のので、約0 でなければならない。		
翻訳ミス	「Voltage-difference : 電圧差」の difference 部分の訳語が欠落している。	該当項目	A

<B. 不要語の混入>

原文中に対応する語が存在しない不要な訳語が機械翻訳結果に混入しているケースを指します。

原 文	Behind the front panel a predetermined distance can be a reflective panel or other surface.		
人手翻訳	前面パネルの後ろの予め定められた 距離 には反射パネルまたはその他の表面があってもよい。		
機械翻訳	フロントパネルの後ろでは、予定された 車間の距離 は、反射的なパネルまたは他の表面であるかもしれない。		
翻訳ミス	原文・人手翻訳に存在しない「車間の」という語が混入している。	該当項目	B

ただし、機械翻訳システムが必要に応じ補足／省略したとみなせる語は、文意に悪影響がなければ不問とします。

[不要語の混入とみなさないケース]

原 文	This invention relates generally to the field of mechanisms used to temporarily connect an implement to a tractor, and more particularly to mechanisms known as three point hitch systems or assemblies.		
人手翻訳	本発明は、器具をトラクタに一時的に接続するために使用される機構の分野全般に関し、より詳細には、3点ヒッチシステム又は3点ヒッチ組立体として知られている機構に関する。		
機械翻訳	本発明は、通常、一時的に道具をトラクタに接続するために用いる機構の分野に関する。より詳しくは、 本願発明は 三点ヒッチ・システムまたはアセンブリとして公知の機構に関する。		
翻訳ミス	機械翻訳結果には、原文中に対応する語が存在しない「本願発明」という語が存在するが、これはエンジンが訳文を2文に分割した際の補足であり、不要語の混入とはみなさない。	該当項目	—

<C. 未知語>

原文の一部が機械翻訳されず原語のまま出力されているケースを指します。ただし、意図的に原語のまま出力していると考えられる語（例：ping packet⇒ping パケット、）は対象外です。

原 文	The side of the container opposite the column 11 is provided with complementary pairs of paddles 21, 22 tiltably mounted to a post 23 at the side of the container.		
人手翻訳	柱11の反対側のコンテナの側面には、角度を変えることができるようにコンテナの側面のポスト23に取り付けられた補完的なパドルの対21、22が設けられている。		
機械翻訳	容器の側面に t i l t a b l y にポスト23に載置されるパドル21、22の相補的な対を、柱11の反対側の容器の側面は、備えている。		
翻訳ミス	Tiltably が未知語となっている。	該当項目	C

<D. 品詞の誤り>

原文中の語が、誤った品詞として翻訳されているケースが対象となります。

原 文	Fig. 4 presents detail of an embodiment of one jaw of the transferring element.		
人手翻訳	図4は、移送構成要素の1つのジョーの実施形態の詳細を 示す 図である。		
機械翻訳	図伝達要素の1顎の形態の4 プレゼント の詳細。		
翻訳ミス	動詞 present（示す）が名詞（プレゼント）と誤って解釈されている。	該当項目	D

原 文	an upper hook and two lower hooks, said lower hooks disposed on two extension legs telescopically retained by two vertical sleeve members;		
人手翻訳	上部フック及び2つの下部フックであって、 前記 下部フックが、2つの垂直スリーブ部材により入れ子状に保持された2つの拡張脚部に配置されており、		
機械翻訳	上部のフックと2下フックは、下のフックは伸縮縦の二重スリーブ部材が保持する2つの拡張脚上に配置されたと 述べた ；		
翻訳ミス	形容詞 said（前記）が動詞（述べた）と品詞誤りされている。	該当項目	D

なお、複数の語から成る複合語（の一部）やイディオムがそのように解釈されていない場合、これも品詞誤りとします。複合語・イディオム単位で1つの語と扱ったうえで本項目にカウントしてください。

[品詞の誤り（複合語）]

原 文	[0024] A heat shield 7 is interposed between the upper face of heating body 3 and housing 1, the heat shield being maintained on the heating body 3 by fastening means.		
人手翻訳	【0024】熱シールド7は 加熱ボディー 3の上面とハウジング1との間に挿入され、該熱シールドは固定手段によって加熱ボディー3上に維持される。		
機械翻訳	【0024】熱シールド7は 本体 3と、ハウジング1を 加熱する の上面との間に介在され、熱シールドは締結手段によって発熱体3上に維持されている。		
翻訳ミス	Heating body（加熱ボディー）が複合語として扱われず、heating が誤って動詞（加熱する）に訳されている。こうした場合、heating body の単位で品詞誤りとみなせばよい。	該当項目	D

添付資料 7-5

また、自動詞と他動詞の誤りは、本項目ではなく「F. 語の誤訳」に該当するものとします。

原 文	None of the outer and inner flanges 723a, 723b can overlap one another when the inner layer 719b is disposed in the outer layer 719a.		
人手翻訳	内側層 7 1 9 b が外側層 7 1 9 a 内に配置される際に、どの外側及び内側フランジ 7 2 3 a、7 2 3 b も互いに 重なり合 わないようにすることができる。		
機械翻訳	内部層 719b が外層 719a に配置されるとき、外側そして内側のフランジ 723a、723b のいずれも、互いに 重ね合わせる ことができない。		
翻訳ミス	原文の overlap は自動詞（重なり合う）で訳すべきだが、機械翻訳結果は他動詞（重ね合わせる）で訳されている。このような場合は、品詞の誤り（D）ではなく語の誤訳（F）にカウントする。	該当項目	F

<E. 係り受け誤り>

機械翻訳結果における各語間の係り受けが正しくない場合、及び、文法規則に反するような語の配置については、この項目にカウントします。カウントは、係り受け関係単位で行います。

原 文	Not infrequently the pulley must be removed from the pin by means of a very quick and simple operation, for example so as to replace it with a different pulley.		
人手翻訳	例えばそれを異なる滑車に置き換えるために、非常に急速で簡単な操作によって該ピンから滑車を取り外さなくてはならないことは 稀ではない 。		
機械翻訳	まれに 、例えばそれを異なるプーリに置き換えるように、プーリは、あまり速いおよび簡単な運営によって、ピンから取り除かれては ならない 。		
翻訳ミス	①「Not infrequently」は「時々」の意味だが、Not が infrequently ではなく must be removed に係り、誤った意味になっている。	該当項目	E

原 文	advantageously, the core 5 is made of nylon, and the layer 6 of conductive metal is for example made of silver.		
人手翻訳	有利には、コア 5 はナイロンでできており、導電性金属の層 6 は、 例えば 銀でできている。		
機械翻訳	有利に、コア 5 はナイロンからなる。また、導電性金属の層 6 は銀からなって、 例えば ある。		
翻訳ミス	「for example：例えば」は「made：なる」に係るべきものだが、そうっていない(「is：ある」に係っている)。	該当項目	E

[係り受け誤りと語の誤訳の区別]

要素の組合せは正しいが、その関係性が誤っている、もしくは助詞の欠落等により関係性が不明瞭な場合も「係り受け誤り」の対象とします。

原 文	The first and second ear straps are bonded to the mask body by a first bonding part and a second bonding part , respectively, which are located away from each other in a first direction respectively.		
人手翻訳	第1及び第2の耳掛け部は、それぞれ第1の方向に沿って離れている 第1の接合部および第2の接合部によって マスク本体部に接合されている。		
機械翻訳	第1及び第2の耳のストラップは、 第1の接合部と第2の接合部は 、それぞれ、第1方向に互いに離隔されるマスク本体に接合されている。		
翻訳ミス	「by a first bonding part～：第1の接合部～」は「bonded to：接合されている」に係るものであり、機械翻訳結果でもその点は正しいが、byを「によって」ではなく「は」と訳しているため、関係性が正しくない。この翻訳ミスは「byの誤訳（F）」と見なすことも可能だが、より上位の本項目にカウントする。	該当項目	E

[係り受け誤りとみなさない例]

例えば能動文から受動文への言い換えなどによって主語・動詞・目的語等の要素が入れ替わり、結果として原文と異なる係り受けとなっている場合でも、原文における各要素の相互関係が正しく把握できるような係り受けであれば、係り受け誤りには該当しません。

原 文	[0021] The strength of the flexible hose can be increased if the pressure support elements comprise a metallic wire.		
人手翻訳	〔0021〕可撓性ホースの 強度は 、圧力支持部品が金属ワイヤーを含む場合に 増加される 。		
機械翻訳	[0021] 圧力支持体素子をメタリックワイヤから構成する場合、可撓性ホースの 強度を増加させることができる 。		
翻訳ミス	原文では主語である strength が機械翻訳結果では目的語となっているが、意味するところは同じであるため、係り受け誤りとはみなさない（pressure support elements についても同様）。	該当項目	—

また、機械翻訳エンジンによっては、係り受け関係を本文中で表現せず、カッコ書きを用いて表現する場合があります。このような場合でも、係り受け関係が正しく把握される限り、カッコ書きを用いたことを理由に翻訳ミスとみなす必要はありません。

原 文	It is preferred that the container is of frame construction, preferably metallic frame construction.		
人手翻訳	コンテナはフレーム構造、好ましくは金属フレーム構造であると好ましい。		
機械翻訳	コンテナがフレーム構造（好ましくは金属的なフレーム構造）であることが好ましい。		
翻訳ミス	Preferably 以下の frame construction への係り受けが、本文ではなくカッコ書きで表現されているが、正しい係り受け関係が維持されているため翻訳ミスとはみなさない。	該当項目	—

※ 不必要なカッコ書きの多用は、「(1) 流暢さの評価」の観点において減点対象となります。

<F. 語の誤訳>

原文中の語が不適切な訳語（同義語として一般的に用いられる語の範疇を外れた語）に訳されている場合、この項目にカウントします。カウントは語単位が原則ですが、複合名詞（例：liquid crystal display）やイディオム（例：among others）等は、全体を1語として扱うものとします。

一文で同じ語が複数回誤訳されている場合は、1回のみカウントしてください。

なお、「(3) 重要技術用語の翻訳精度の評価」にて対象とした「重要技術用語」も、再度評価対象となります。原則、「(3) 重要技術用語の翻訳精度の評価」にて B～C となった語はこの項目に該当するものとします（※品詞誤りを除く）。

原 文	The arm usually forms an angle of 90 with the base .		
人手翻訳	アームは通常、 基部 と90度の角度を形成する。		
機械翻訳	アームは、通常90の角度を 塩基 により形成する。		
翻訳ミス	「Base：基部」が「塩基」と、異なる技術分野の訳語に誤訳されている。	該当項目	F

原 文	The contact member advantageously comprises a flexible or resilient contact layer for contacting the adjacent item in the container.		
人手翻訳	接触部材は、コンテナの中で隣り合う 品物 と接触するための可撓性接触層または弾力的な接触層を含むと有利である。		
機械翻訳	接触部材は、有利には、容器内の隣接する 項目 と接触するための可撓性又は弾性接触層を含む。		
翻訳ミス	Item は「項目」ではなく「品物」等の意味。	該当項目	F

原 文	A further slide bar 62 is fixed at the bottom of the container adjacent openable door panel 5.		
人手翻訳	コンテナの基部において、開くことができる扉の板 5 と隣り合って別の スライドバー 6 2 が固定されている。		
機械翻訳	更なる すべり棒 6 2 は、開けられるドアパネル 5 に隣接してコンテナの底で修正される。		
翻訳ミス	「Slide bar : すべり棒」は技術的に同義かどうか微妙であるが、一般的に使われる表現ではないため、誤訳と見なす。	該当項目	F

<G. その他の翻訳ミス>

上記 A～F のいずれの項目にも該当しない翻訳ミスは、ここにカウントします。具体例として、例えば以下のようなケースが考えられます。

原 文	For example in the embodiment shown in the drawings, doors 5 and 7 pivot on respective support posts to effectively open respective corners of the container, thereby to enhance access to the ends of the container for loading or unloading of the container.		
人手翻訳	たとえば、図面に示された実施態様において、扉 5 および 7 はそれぞれの支柱を軸として回転してコンテナのそれぞれの角を効果的に開き、それによって、コンテナの荷積みまたは荷降ろしのためのコンテナの端への出し入れを容易にする。		
機械翻訳	ドローイングの実施例で ネットワーク・イメージ用のネットワーク共有を作成する場合は、ルートのインストール先の UNC パスを半角 39 文字未満にしてください 、コンテナの事実上開いているそれぞれのコーナへのそれぞれの支柱のドア 5 と 7 旋回軸; その結果、コンテナの装填または取り降ろしのためのコンテナの端部への機能アップアクセスに。		
翻訳ミス	機械翻訳結果中に原文と無関係な文が混入しており、内容的に「B. 不要語の混入」の範疇を超えているため、こちらに分類した。	該当項目	G

原 文	To drive two toggle levers (29, 29') of a press,...		
人手翻訳	プレス機の 2 つのトグルレバー (29、29') を駆動するために、...		
機械翻訳	プレスの 2 つのトグルレバー (2 9' および 2 9) を駆動するために、		
翻訳ミス	カッコ中の付番 29 と 29' の順序が逆になっている。	該当項目	G

※注記

翻訳ミスが A～F のどの項目に該当するか特定困難な場合は、最も可能性の高いと思われる項目にカウントしてください。それも不明な場合は、「G. その他の翻訳ミス」にカウントしてください。

翻訳結果が存在しない、または翻訳結果に重大なエラーがあり評価困難な場合は、「G. その他の翻訳ミス」に 999 をセットしてください。

<Z. 任意項目>

検出した翻訳ミスは、まずは必須項目 A～G のいずれかにカウントしたのち、これら 7 項目とは別の切り口から設けた本項目とも照合し、該当する基準の全てにカウントします。本項目の基準は以下の 12 種です。

Z-00・訳語のゆれ

Z-01・不適切な文の分断

Z-02・主語・述語・目的語（他動詞の場合）の相互関係に関する翻訳ミス

Z-03・並列表現に関する翻訳ミス

Z-04・発明の名称など、名詞句を文として扱った翻訳ミス

Z-05・特許特有の用語や表現に関する翻訳ミス

Z-06・特許特有の記号や文献番号表記、付番に関する翻訳ミス

Z-07・箇条書き表現に関する翻訳ミス

Z-08・原文のカッコ書きに起因する翻訳ミス

Z-09・数式、化学式に関する翻訳ミス

Z-10・数字と単位に関する翻訳ミス

Z-11・数値範囲に関する翻訳ミス

[Z-00. 訳語のゆれ]

原文中に同じ用語が複数回出現し、機械翻訳結果においてそれらが異なる訳語で訳されていると、読解に悪影響を及ぼす場合があります。項目 Z-00 では、原文中に同一の用語が複数回出現し、それらの訳語が統一されていないことで読解に悪影響を及ぼす場合にカウントします。ただし、冠詞や前置詞をはじめ、訳語が統一されている必要のない語は対象外とします。また、各訳語の適・不適を考慮する必要はありません。カウントは（出現回数単位でなく）原語単位で行います。

※本項目のみ、【必須項目】A～F のいずれにも該当しない場合があります（訳語は異なるが、どちらも適訳である場合）。この場合、【必須項目】は「G. その他の翻訳ミス」にカウントしてください。なお、A～F のいずれかに該当する場合は、G へのカウントは不要です。

原 文	[0025] Housing 1 contains a water reservoir 8 being composed of an upper part 9 and a lower part which are assembled on one another, heat shield 7 defining the lower part of water reservoir 8.		
人手翻訳	【 0 0 2 5 】ハウジング 1 は、互いに組み立てられる上側部分 9 および下側部分から構成される 水リザーバ 8 を包含し、熱シールド 7 が 水リザーバ 8 の下側部分を画定する。		
機械翻訳	[0025]はハウジング 1 上部 9 と互いに組み立てられ下部で構成されている 貯水槽 8 が含まれている、熱シールド 7 は 水の 下部 貯水池 8 を定義する。		
翻訳ミス	原文中 2 度出現する water reservoir について、機械翻訳結果では 1 回目が「貯水槽」、2 回目が「水の／貯水池」と異なる訳語に訳されており、訳語のゆれに相当。 ※2 回目の water reservoir は複合語として扱われていないため必須チェック項目の D にも該当する。	該当項目	Z-00

同一の語であっても、一方が単語、他方が複合語の一部であったり、それぞれが異なる複合語に含まれていたりする場合があります。このような場合は、同一の訳語に訳されないと読解に支障をきたすかどうかで判断してください。

原 文	It will be appreciated that a given application service is dedicated to a predetermined application .		
人手翻訳	特定の アプリケーション ・サービスは、所定の アプリケーション に専用されることが理解されよう。		
機械翻訳	特定の アプリケーション ・サービスが予定された 塗布 に捧げられるのが理解される。		
翻訳ミス	原文に 2 度出現する application は、1 回目は application service という複合語、2 回目は application という単語であるが、文意から同一の事項を指しており、同一の訳語に訳すべきものである。 ※「application : 塗布」は「F. 語の誤訳」にも該当する。	該当項目	Z-00

[Z-01. 不適切な文の分断]

例えば入力文が長すぎて処理できない場合や、記号等に用いられたピリオドを誤って文末と判断してしまった場合など、不適切な文の分断が検出された場合、ここにカウントします。

ただし、例えば要素列挙時や、適切な箇所での意図的な分割など、正当な理由が認められる場合はカウントは不要です。

原 文	[0014] FIG. 3 shows a cross-sectional view of the fixing device according to FIG. 1 along the cross-sectional plane I-I of FIG. 1 .		
人手翻訳	【0014】図3は、 図1 の横断平面 I-I に沿う 図1 に従う固定装置の横断面図を示す。		
機械翻訳	【0014】 図 。図3は、図1に係る定着装置の断面図を示す。 断面平面図の I-I に沿って 1。		
翻訳ミス	FIG. 1 のピリオドが文末と解釈され、文が分断されている。 2 箇所において発生しているため、Z-01 は 2 回カウントする。 ※FIG.1 は一語とみなされ、そう扱われていないため項目 D にも該当する。また、後述の Z-06（特許特有の記号に関する翻訳ミス）にもカウントする。	該当項目	Z-01×2

[Z-02. 主語・述語・目的語（他動詞の場合）の相互関係に関する翻訳ミス]

原文における骨格構造、すなわち主語・述語・目的語（他動詞の場合）の相互関係が正しく保たれていない場合、ここにカウントします。

原 文	The ping packets is expected to return a response within the TTL = 1.		
人手翻訳	ピング・パケットは、TTL = 1 内に応答を返すことが期待される。		
機械翻訳	ピング・パケット TTL = 1 の中で応答を返す予想する。		
翻訳ミス	原文の骨格構造は「主語＝ピングパケット、動詞＝期待される」だが、機械翻訳結果は、ピングパケットが主語かどうか不明瞭であり、かつ動詞「予想する」が（意味が誤っていることを不問としても）能動態となっているため、骨格構造を維持できていない。 ※A～H 項目においては「ping packet と is expected」の係り受け誤り（E）と、expect の誤訳（F）の 2 つの翻訳ミスとしてカウントが必要だが、Z-02 へのカウントは 1 回でよい。	該当項目	Z-02

It is preferred that XXX など、It や This 等の代名詞が便宜的な主語として用いられている場合は、実質的な主語（XXX）を評価対象としてください。

原 文	It is preferred that one or more contact members are height adjustable with respect to the container.		
人手翻訳	1 つ以上の接触部材がコンテナに対して高さ調節可能であると好ましい。		
機械翻訳	それは 、一つ以上の接触部材は、容器に対して高さ調節可能であることが好ましい。		
翻訳ミス	It は便宜的な主語であって翻訳文は that 以下を主語として扱うべきであるが、機械翻訳結果では it をそのまま主語として扱い「それは」と訳しているため、本項目に該当するとみなした。 ※翻訳すべきでない It を訳しているという翻訳ミスは、「G. その他の翻訳ミス」とした。	該当項目	Z-02

なお、例えば能動文から受動文への言い換えなどにより主語・動詞・目的語が入れ替わるような場合も、相互関係が正しく把握されていれば問題ありません。

原 文	Energy absorption means (21) is provided at each end for absorbing energy transmitted by the ropes from the impacting vehicle.		
人手翻訳	衝突車両からロープによって伝達されるエネルギーを吸収するために、各端部には エネルギー吸収手段（21）が設けられる 。		
機械翻訳	各端のときにロープで衝撃付与乗り物からエネルギー透過を吸収するのに エネルギー吸収手段(21)を設ける 。		
翻訳ミス	原文は「エネルギー吸収手段が設けられる」という受動文。機械翻訳結果ではこれを「エネルギー吸収手段を設ける」と能動態に言い換えているが、主語と動詞の関係性は維持されているので翻訳ミスとはみなさない。	該当項目	—

また、主語や動詞が誤訳されて正しい意味が把握できなくても、文の骨格構造が維持されていれば本項目には該当しません。

原 文	The tongue engages inside a corresponding seat of the pulley once the latter has been inserted on the pinion.		
人手翻訳	該 トング は、滑車がピニオンに挿入されると滑車の対応する シート の内側に 係合する 。		
機械翻訳	後者はピニオンに挿入された後に 舌 が滑車の対応 座席 の内側 を行っています 。		
翻訳ミス	tongue が「舌」、engage が「行っている」と誤訳されているため文意は把握困難だが、文の骨格構造は正しいため本項目には該当しない。	該当項目	—

本項目は原文が文であるもののみを対象とします。従って、「発明の名称」や「特許請求の範囲」など、対象文が名詞句である場合は対象外となります。「特許請求の範囲」などにおいて、こうした名詞句を機械翻訳エンジンが文形式で翻訳することがありますが、この場合も対象外です。

原 文	4. Device (10) according to claim 3, wherein said element (60) which can be operated by the user is slidable along said cavity (14).		
人手翻訳	4. 請求項3に記載の装置（10）において、ユーザにより操作され得る前記エレメント（60）は前記空洞（14）に沿ってスライド可能である、装置。		
機械翻訳	4.請求項3に記載のデバイス（10）において、ユーザにより作動されることができる前記素子（60）は、前記空腔（14）に沿って摺動可能である。		
翻訳ミス	機械翻訳結果は意図的に文形式で翻訳されており主語、述語が存在するが、原文が名詞句であるため、本項目の対象としない。	該当項目	—

[Z-03. 並列表現に関する翻訳ミス]

原文に含まれる並列関係が機械翻訳結果において正しく認識されず、その構造が崩れていたり、双方に係るべき修飾語が一方のみに係っていたりする場合にカウントします。並列表現は、付番の並列や語の並列から節の並列、文の並列まで、種類を問わず対象とします。

原 文	A second duct includes a damper or valve which may be detachable.		
人手翻訳	第2のダクトは、取り外し可能な ダンパー または バルブ を備える。		
機械翻訳	別のダクトは、 ダンパー または取り外せる バルブ を有する。		
翻訳ミス	damper と valve が並列関係にあり、双方を which 以下が修飾しているが、機械翻訳ではそう見なされず、which が valve のみを修飾している。 ※which 節の係り受け誤り（E）にも該当する。	該当項目	Z-03

なお、本項目の評価観点は「並列関係が保たれているか」という一点のみであり、それさえ保たれていれば、他種の翻訳ミス（例えば語の誤訳）が発生していても本項目のカウント対象とはしません。

原 文	A second duct includes a damper or valve which may be detachable.		
人手翻訳	第2のダクトは、取り外し可能な ダンパー または バルブ を備える。		
機械翻訳	第2ダクトは、取り外せる ピアノダンパー または バルブ を備える。		
翻訳ミス	Damper がピアノダンパーと誤訳されているが、valve との並列関係は which 節の係り受けを含め保たれているので、本項目には該当しない。	該当項目	—

並列表現において繰り返し部分が省略されるケースがあり、これに適切に対応できているかどうか本項目において評価観点となります。

原 文	As an additive and modificative process , after a trench portion is formed in the inorganic insulating film, ...		
人手翻訳	追加的、修正的プロセス として、無機絶縁膜にトレンチ部を形成した後、トレンチ部の内壁面に、付着膜として、...		
機械翻訳	添加剤および変更するプロセス として、溝部分が例えば無機絶縁膜(CVD 薄膜)の中で形成された後、...		
翻訳ミス	Additive と modificative はともに process を修飾する形容詞であり、additive の後の process が省略されている。機械翻訳結果は語の並列関係自体は正しいが、この省略に対応できずに additive を「添加剤」と名詞解釈している。 ※additive は品詞誤り (D) にも該当。	該当項目	Z-03

繰り返し語の省略は、動詞や助動詞等においても行われる場合があります。

原 文	At the unloading location, the brake system can be disengaged and the paddles released .		
人手翻訳	荷降ろしの場所において、ブレーキシステムの係合を解き、パドルを 離すことができる 。		
機械翻訳	取り降ろし所在地では、ブレーキシステムは、解放されていて リリースされた パドルであるかもしれない。		
翻訳ミス	末尾の the paddles released は正しくは the paddles can be released であり can be が省略されているが、そう解釈されず、released が後置修飾の形容詞と判断されている。 ※released の品詞誤り (D) にもカウント。なお上例では disengaged も形容詞に誤解釈されている。	該当項目	Z-03

また、複数の並列関係が複合するケースもあります。この場合、個々の並列関係ごとに評価対象とします。

原 文	In addition, solar and otherwise powered, lighted signs are common.		
人手翻訳	加えて、太陽電池または別様で電力供給される点灯式標識も一般的である。		
機械翻訳	太陽で、動力がそうでなければ供給されて、さらに、明かりのついたサインは一般的である。		
翻訳ミス	並列表現は①solar と otherwise powered、そして②solar and otherwise powered と lighted の 2 箇所。①②とも sign を修飾しているが、機械翻訳結果では、どちらも正しく把握できないため①、②とも本項目に該当。従って Z-03 は 2 カウントとなる。 ※他に、otherwise が品詞誤り (D)、In addition と solar がそれぞれ係り受け誤り (E)、powered が語の誤訳 (F) にそれぞれ該当する。	該当項目	Z-03×2

[Z-04. 発明の名称など、名詞句を文として扱った翻訳ミス]

原文が名詞句である場合において、機械翻訳結果にてこれを誤って文と解釈している場合、ここにカウントします。通常の文においても、文の一部を構成する名詞句が文として誤解釈されていればカウント対象となります。

原 文	TRANSFERRING ELEMENT OF A PUSHER MECHANISM OF A GLASS FORMING MACHINE		
人手翻訳	ガラス成形機のプッシャ機構の 移送構成要素		
機械翻訳	ガラスフォーミングマシンのプッシャー機構の 原理を移す 。		
翻訳ミス	名詞句を文として訳している。 ※品詞誤り (D) にも必ず該当する。	該当項目	Z-03

なお、機械翻訳結果において名詞句が意図的に文として訳されている場合はカウント不要です。

原 文	2. The integrated circuit chip of claim 1, wherein the at least one functional domain includes an RF macro.		
人手翻訳	【請求項 2】前記少なくとも 1 つの機能領域が R F マクロを含む、請求項 1 に記載の集積回路チップ。		
機械翻訳	【請求項 2】請求項 1 に記載の集積回路チップは、少なくとも 1 つの機能性領域は、R F マクロを含むものである。		
翻訳ミス	機械翻訳結果は請求項の典型的な名詞句パターンを意図的に文形式で翻訳していると判断されるため、そのこと自体は翻訳ミスとみなさない。	該当項目	—

[Z-05. 特許特有の用語や表現に関する翻訳ミス]

例えば英語特許文献であれば「detailed description：詳細な説明」や「preferred embodiment：好適な実施形態」、「incorporated by reference：参考として援用された」など、通常の文では減多に用いられないが、特許文献においては頻繁に用いられる用語や表現に関する翻訳ミスをカウントします。技術用語全般は対象となりません。

原 文	One skilled in the art will understand that numerous changes and variations are possible without departing from the spirit of the invention .		
人手翻訳	本発明の趣旨 から逸脱することなく多数の変更および変形が可能であることを 当業者 は理解するだろう。		
機械翻訳	技術に熟練している 1 、多数の変更および変化が 発明のスピリット から外れずに可能であると理解するだろう。		
翻訳ミス	「one skilled in the art：当業者」「spirit of the invention：発明の趣旨」は特許特有の表現であり、機械翻訳結果はこれらを正しく訳せていないため本項目にカウントが必要。 ※どちらも 語の誤訳 (F) にも該当する。	該当項目	Z-05×2

なお、「発明の名称」における代名詞 same や「特許請求の範囲」等における wherein 等、語としては一般語であっても特許文献にて特有、もしくは頻出かつ典型的な用いられ方をする語については、本項目の対象としてください。

[Z-06. 特許特有の記号や番号表記、付番に関する翻訳ミス]

具体的には、図表番号（Fig.1 や FIG1）や文献番号（US.Pat.No.）等の記号や番号表記に関する翻訳ミスをカウントします。下例のような要素への付番も対象となります。

なお、段落番号（[0001]、[0002]）は、箇条書き表現の一種とみなし、「Z-07.箇条書き表現に関する翻訳ミス」にて対象とするため本項目では対象外です。

原 文	The locking teeth 52a,b are connected to the arms 54a,b by means of two inclined segments 62a,b.		
人手翻訳	ロッキング歯 5 2 a、b は 2 つの傾斜セグメント 6 2 a、b によってアーム 5 4 a、b に結合されている。		
機械翻訳	係止歯 52a は、b は 2 つの傾斜部セグメント 62a、b によって腕部 54a、b に接続されている。		
翻訳ミス	Locking teeth の付番 52a,b が正しく解釈されず、b が単独の要素として扱われている。 ※係り受け誤り（E）のほか、a と b は並列表現でもあるため Z-03（並列表現に関する翻訳ミス）へのカウントも必要。	該当項目	Z-06

[Z-07. 箇条書き表現に関する翻訳ミス]

評価対象文が箇条書き文に該当し、その箇条書き表現部分において翻訳ミスが発生している場合、本項目にカウントします。特許文献で多用される段落番号や請求項番号も、箇条書きの一種に含めます。

原 文	[0034] The crystal consequently obtained was cut perpendicularly to the direction of growth.		
人手翻訳	[0 0 3 4] その結果得られた結晶を成長方向に対して垂直に切断した。		
機械翻訳	[0034] は、その結果得られた結晶を成長方向に垂直に切断した。		
翻訳ミス	段落番号[0034]が文の一部として扱われてしまっている。 ※段落番号を名詞と誤解釈しているため、品詞の誤り（D）にも該当。	該当項目	Z-07

注) 請求項番号に関しては、原文の箇条書き番号が一般的な項番「1.」「2.」であるのに対し、人手翻訳が【請求項 1】【請求項 2】となっている場合がありますが、機械翻訳結果においては一般的な項番のままでも可とします（原文が単文単位につき、記載箇所の特定が難しいためです）。

原 文	5. Testing method according to claim 1 wherein the sample of ore to be tested is between 2 and 10 kg in weight.		
人手翻訳	【請求項 5】試験される前記鉱石試料は、重量が 2 k g から 1 0 k g の間である、請求項 1 に記載の試験方法。		
機械翻訳	5. テストされる鉱石のサンプルが重さ 2～10kgであることを特徴とする請求項 1 に記載の試験方法。		
翻訳ミス	箇条書き番号が人手翻訳では【請求項 5】となっているのに対し機械翻訳結果は通常の番号のままだが、原文には忠実であるため問題なし。	該当項目	—

[Z-08. 原文のカッコ書きに起因する翻訳ミス]

評価対象文の原文がカッコ書きを含み、そのカッコ書き表現に起因した翻訳ミスが発生している場合、ここにカウントします。具体的には、カッコで括られた範囲が原文と異なる、カッコ書き部分と他の部分との係り受けの誤り等が該当し、カッコ内の語間の係り受け誤りや用語の誤訳など、カッコ書きとの因果関係が認められない翻訳ミスは対象となりません。

また、原文にないカッコ書き表現が機械翻訳結果にて用いられることがあります。本基準の対象はあくまで原文のカッコ書きに起因する翻訳ミスですので、これらは対象外となります。

原 文	The force 40 exerted by the vertical stream 30 of air is higher than the one in the case of known structures with horizontal blowing as the air stream 30 flows vertically around the product 2 and acts on a larger area 44 of the product 2 (see fig. 7).		
人手翻訳	空気の鉛直な流れ 3 0 によって働かされる力 4 0 は、水平方向の吹き込みを伴う公知の構造の場合の力よりも高く、これは気流 3 0 が製品 2 の周りを鉛直に流れて製品 2 のより大きな範囲 4 4 に作用するためである (図 7 を参照)。		
機械翻訳	空気の垂直流 30 によって及ぼされる力 40 は、水平の (図 30 を参照して、オリジナル 2 のより大きな領域 44 で製品 2 作用周りに上下方向に流れる空気流として送風で知られている構造の場合、1 つ以上である 7)。		
翻訳ミス	カッコ書き表現 (see fig. 7) の単位が維持されず、本文の内容がカッコ中に混入してしまっている。 ※係り受け誤り (E) と特許特有記号 (Z-06) にもカウントする。	該当項目	Z-08

[Z-09. 数式、化学式に関する翻訳ミス]

評価対象文が数式や化学式を含み、その数式や化学式において翻訳ミス（式が途中で分断されている等）が発生している場合、ここにカウントします。

原 文	Given a triangle with vertices A, B, and C, the pixel intensity at point D can be obtained as: $D = A \cdot w_A / T_{sub.abc} + B \cdot w_B / T_{sub.abc} + C \cdot w_C / T_{sub.abc}$ (2)		
人手翻訳	頂点A、B、およびCを有する三角形を考慮すると、点Dでの画素輝度は、以下の数式3で取得されることが可能である。 $D = A \times w_A / T_{abc} + B \times w_B / T_{abc} + C \times w_C / T_{abc}$		
機械翻訳	頂点A、B、およびC、点Dでのピクセル強度を得ることができるとともに、三角形所与： $D = A \cdot w_A / T_{sub.abc} + B \cdot w_B / T_{sub.abc} + C \cdot w_C / T_{sub.abc}$ (2)		
翻訳ミス	数式部分がそう認識されず、助詞が混入している。 ※助詞は不要語の混入ともみなせるが、数式は一語として扱うべきところ通常の文として扱った結果であるため、品詞の誤り（D）とするのがより適切。	該当項目	Z-09

[Z-10. 数値単位に関する翻訳ミス]

数字とそれに付随する単位（例：meter、minute、kg、% by weight）に関する翻訳ミスをカウントします。数字と単位が分断されている場合だけでなく、単位が誤訳されている場合も対象となります。

原 文	Wafers of such seed crystal grown to the approximate level of 3 inches are now available on the market.		
人手翻訳	こうして約3インチレベルに成長させた種結晶のウェーハが現在市販されている。		
機械翻訳	3 つ の大体のレベルまで育てられたそのような種晶のウェファースは現在、 何インチ も市販で入手できる。		
翻訳ミス	原文の 3 inches が数字＋単位として扱われていない。 ※3 inches で一語とみなすべきであり、これが一語として扱われていないため品詞誤り（D）に該当。	該当項目	Z-10

[Z-11. 数値範囲に関する翻訳ミス]

「from NNN to NNN」や「NNN 未満」等、数値範囲を正しく処理できていない場合、ここにカウントします。

原 文	Advantageously, the fraction of weight of the electrically conductive filament 4 compared with the weight of the yarn 1 according to the present implementation method is between 5% and 50%, and preferably between 30% and 40% , for example approximately 35%.		
人手翻訳	有利には、本実施方法による糸 1 の重量に対する導電性フィラメント 4 の重量の割合は、5 %から 5 0 %、好ましくは 3 0 %から 4 0 % 、例えば約 3 5 %である。		
機械翻訳	有利には、糸 1 の重量と比較して電氣的に導電性フィラメント 4 の重量の本実装方法に係る分率は 5 %～5 0 %、 3 0 %、4 0 % 、例えば、約 3 5 %の間であることが好ましい。		
翻訳ミス	数値範囲は「 between 5% and 50% 」と「 between 30% and 40% 」の 2 箇所あるが、機械翻訳では後者が数値範囲であることが表現できていない。 ※係り受け誤り（E）に該当する。	該当項目	Z-11

なお、数値範囲は数値単位表現を伴うことが多いですが、本項目で対象とするのは数値範囲の部分のみであり、数値単位に関する翻訳ミスは項目 Z-10（「数字と単位に関する翻訳ミス」）のみにカウントしてください。

原 文	The recommended blink rate is around 55-60 times a minute.		
人手翻訳	推奨される点滅速度は 1 分間に約 5 5 ～ 6 0 回である。		
機械翻訳	推奨点滅速度は 55-60 倍分 程度である。		
翻訳ミス	機械翻訳結果は数値範囲「 55-60 」を正しく訳しているので本項目の対象とはならないが、数値単位 times を「倍」と誤訳しているので Z-10 にはカウントが必要。 ※語の誤訳（F）にも該当する。	該当項目	Z-10

以 上

添付資料 4-2 : 「チェックリストに基づく評価（英日項目）」詳細説明

本資料では、「チェックリストに基づく評価」の翻訳言語別任意項目のうち、英日機械翻訳に関する下記 16 項目について、各基準に相当する機械翻訳結果の実例ならびにその評価根拠を示します。

- EJ-01・文中のピリオドを適切に処理していない
- EJ-02・「:」や「;」を用いた文を適切に処理していない
- EJ-03・重要な単数・複数情報が維持されていない
- EJ-04・単数形・複数形併記表現を適切に処理していない
- EJ-05・序数詞表現が正しく解釈されていない
- EJ-06・慣用句の中間への語の挿入を適切に処理していない
- EJ-07・倒置文を適切に翻訳していない
- EJ-08・名詞と代名詞の関係を適切に表現できていない
- EJ-09・慣用表現とみなすべき場合／そうでない場合を適切に翻訳できていない
- EJ-10・前置詞に対し適切な訳語を選択できていない
- EJ-11・接頭辞、接尾辞を伴う語を適切に処理していない
- EJ-12・スラッシュを用いた並列表現を適切に処理していない
- EJ-13・therein, therefor, wherein, 等を適切に処理していない
- EJ-14・代名詞 same、the same を適切に処理していない（【発明の名称】）
- EJ-15・小文字を含む大文字表記を適切に処理していない（【発明の名称】）
- EJ-16・クレーム引用に関する表現を適切に翻訳していない（【特許請求の範囲】）

<EJ-01. 文中のピリオドを適切に処理していない>

文中にてピリオドが記号の一部などで使われた際、これが誤って文区切りとして扱われる等、適切に処理できていないものが該当します。

原 文	Some signs are hand-held such as the sign taught by U.S. Pat. No. 6,409,358 that uses a flashlight to internally illuminate it.		
人手翻訳	標識のあるものは手持ち型であり、たとえば 米国特許第 6, 4 0 9, 3 5 8 号 により教示される標識は内部を照らすために懐中電灯を用いる。		
機械翻訳	いくつかの兆候は、 米国特許 により教示印として手持ちされています。内部でそれを照らすために懐中電灯を使用して 番号 6409358 。		
翻訳ミス	文献番号表記 U.S. Pat. No.にはピリオドが4つ含まれる。機械翻訳結果では、このうち Pat の後のピリオドを文区切りと見なしてしまっている。U、S、No の後のピリオドは正しく処理されているので不問だが、もしこれらも文区切りとして扱われていたら、ピリオド単位でカウントする。 ※US. Pat. No.が一語として扱われていない (D)、不適切な文の分断 (Z-01)、特許特有の記号に関する翻訳ミス (Z-06) にも該当。	該当項目	EJ-01

<EJ-02. 「:」や「;」を用いた文を適切に処理していない>

英語特許文献では、クレーム部等にて、構成要素を一文中で「:」や「;」で区切って列挙する文が多用されます。文に含まれる「:」や「;」に適切に対応できていない場合、ここに該当します。

原 文	7. Electrically conductive textile yarn (1) according to any one of the preceding claims, wherein said second element (3) consists substantially of fibres selected from the group comprising: aramide fibres, meta-aramide fibres, Lenzing FR(R).		
人手翻訳	7. 前記第2の要素(3)が、 以下の 群から選択されるファイバーから実質的になる、先行請求項のいずれか一項に記載の導電性繊維糸(1): アラミドファイバー、メタアラミドファイバー、L e n z i n g F R (登録商標)。		
機械翻訳	7。請求項1に記載の導電性の繊維糸(1)、前記第2の要素(3)が実質的に含む群から選択される繊維 からなる 前記: アラミド繊維、メタ・アラミド繊維、レンチング FR (R)。		
翻訳ミス	原文は「comprising:」の後に要素が列挙されており、「:」で文を分割する場合はそれを踏まえた対応をする必要があるが、翻訳結果は「:」前後の関係性を表現できていない。 ※comprising が group に係っていないため、係り受け誤り (E) に該当する。	該当項目	EJ-02

＜EJ-03. 重要な単数・複数情報が維持されていない＞

英語には名詞に単数形・複数形の区別がありますが、日本語にはありません。このため通常は単数形・複数形とも同じ訳語で訳されます。しかしながら、英文のなかには単数形と複数形の使い分けに大きな意味を持たせているケースがあります。このような文において、そのニュアンスが適切に表現できていない場合、ここに該当します。

※ 本項目に該当する翻訳ミスは、必須チェック項目のA～Fのいずれにも該当しないため、必須チェック項目では「G. その他の翻訳ミス」にカウントします。

原 文	In the embodiment shown the arrangement includes upper and lower pairs of paddles 17, 18 mounted tiltably to the column 11.		
人手翻訳	図示された実施態様において、本装置は、角度を変えることができるように柱11に取り付けられた パドル17、18の上下2対 を含む。		
機械翻訳	… パドル17、18の上下の対 …		
翻訳ミス	原文は pairs と複数形にすることによって、パドル17、18の対が上下で2対であるということが明白だが、翻訳結果では、そのニュアンスが希薄である。 ※「pairs」を「対」と訳すこと自体は「F. 語の誤訳」とは言えないため、必須項目は「G. その他の翻訳ミス」にカウントする。	該当項目	EJ-03

なお、この基準の対象は、「単数形と複数形の使い分けに大きな意味を持たせている表現」に限ります。通常の文において単数・複数を区別して訳さないのは日本語の性質によるものであり、不問とします。

評価作業においては、基準ごとに評価対象とすべき文が明示されますので、それに従ってください。

<EJ-04. 単数形・複数形併記表現を適切に処理していない>

英文では、要素が「単数および／または複数」である場合、これを単数形・複数形併記表現（例：an element and/or elements）やカッコ補記表現（例：element(s)）で示す場合があります。機械翻訳結果がこうした表現に適切に対応できていない場合、この基準にカウントします。

※ 本項目に該当する翻訳ミスは、必須チェック項目の A～F のいずれにも該当しないため、必須チェック項目では「G. その他の翻訳ミス」にカウントします。

原 文	a container according to any preceding claim, wherein a contact member or members is mounted to the container		
人手翻訳	前記 1 つ以上の接触部材 は、前記コンテナに取り付けられている、先行請求項のいずれか 1 項に記載のコンテナ。		
機械翻訳	前記いずれかの請求項記載のコンテナであり、 接触部材又は部材 がコンテナに装着されているもの		
翻訳ミス	原文は「1 つまたは複数の接触部材」というニュアンスだが、翻訳結果では意味のない繰り返し表現となってしまう。 ※members を「部材」と訳すこと自体は「F. 語の誤訳」とは言えないため、必須項目は「G. その他の翻訳ミス」にカウントする。	該当項目	EJ-04

<EJ-05. 序数詞表現が正しく解釈されていない>

冠詞に関する特許英文頻出表現として、不定冠詞＋序数詞表現（a first XXX, a second XXX）や、無冠詞＋序数詞表現が多用されることが挙げられます。こういった原因により、序数詞表現が序数詞と解釈されていない場合、この基準に該当します。

原 文	Heat shield 7 to this end has a first channel 71 and a second annular channel 72 receiving, with interposition of a sealing strip, the walls of upper part 9.		
人手翻訳	熱シールド7は、この目的のために、 第1 チャンネル7 1 と、シーリングストリップを介在させて上側部分9の壁を受け入れる 第2 の環状チャンネル7 2 とを有する。		
機械翻訳	第1 は、このために熱シールド7で71と 2分の1 の環状流路72受信にチャンネルを開設する、シールストリップの挿入物で、上部9の土手		
翻訳ミス	「第2の」と訳すべき a second が「2分の1の」と誤訳されている。 second は「秒」と誤訳されることも多い。ここでは a first の係り受けも誤っている。 ※「a second : 2分の1の」は形容詞として annular channel に係っているため品詞誤りや係り受け誤りには該当せず、語の誤訳（F）となる。	該当項目	EJ-05

なお、本項目は序数詞がそのように解釈されていない場合のみにカウントします。例えば **first** が「最初の」、**second** が「次の」等と訳されており、それが文脈上不適訳であったとしても、序数詞として扱われている限り、本項目のカウントは不要です。

<EJ-06. 慣用表現への語の挿入を適切に処理していない>

英文においては、複数の語から成る慣用表現の中間に語が挿入されることがよくあります。英日機械翻訳結果において、こうした語の挿入により、慣用表現をそれと認識できず誤訳されたと見なされる場合、この基準に該当します。

原 文	For example, sacrificial anode layers may be placed in electrical contact with metal surfaces to be protected from corrosion.		
人手翻訳	例えば、犠牲アノード層は、腐食から保護すべき金属表面と 電氣的に接触して 配置することができます。		
機械翻訳	例えば、犠牲陽極層は腐食から保護される金属表面を 備えた電気接点に 置かれるかもしれません。		
翻訳ミス	in contact with は「～に接して」という慣用表現だが、本例ではここに electrical が挿入されている。機械翻訳結果は、これを慣用表現への語の挿入と認識できず、個々の語として扱ってしまっている。 ※慣用表現をそう扱っていないため品詞誤り (D) となる。	該当項目	EJ-06

<EJ-07. 倒置文を適切に翻訳していない>

倒置文の文構造を正しく認識していないと思われる場合、ここにカウントします。文全体が対象となるため、1 文につきカウントは 1 回のみとなります。

原 文	Spanning the width at the base of container 1 are three slide bars 9,10, 61 each having low friction hard upper surfaces.		
人手翻訳	コンテナ 1 の基部において、低摩擦硬質上面をそれぞれ有する 3 つのスライドバー 9、10、61 が幅の方向に横切っている。		
機械翻訳	コンテナ 1 の基礎で幅を測ることは、低摩擦を各々困難にしておく 3 本のスライドバー 9、10、61 である、上部の表面。		
翻訳ミス	原文は are の前後が入れ替わった倒置文であるが、うまく対応できていない。 ※are の前後の係り受け誤り (E) となる。その他、個々の要素ごとに別途カウントが必要 (例えば low friction hard upper surface の品詞誤り)。	該当項目	EJ-07

<EJ-08. 名詞と代名詞／省略表現の関係を適切に表現できていない>

同一文中にて既出の名詞が再度出現する場合、代名詞や省略表現（例：machine translation system であれば the system 等）が用いられることが多いですが、文の構造によっては、機械翻訳結果にて両者の順序が入れ替わり、関係性がわかりにくくなる場合があります。こうした問題が生じている場合、ここにカウントします。

※ 本項目に該当する翻訳ミスは、必須チェック項目の A～F のいずれにも該当しないため、必須チェック項目では「G. その他の翻訳ミス」にカウントします。

原 文	Paddles 17, 18 tiltably supported by the restraint 11 may engage the items to secure them .		
人手翻訳	角度を変えられることができるように固定具 11 によって支持されたパドル 17、18 が 品物 と係合して 品物 を確保することができる。		
機械翻訳	拘束具 11 により傾斜して支持されるパドル 17、18 は、 それら を固定するために 物品 を拘束してもよい。		
翻訳ミス	原文末尾の them は items を指すが、翻訳結果では them のほうが先に出現してしまうため、語順的にはパドル 17,18 を参照しているように受け取られてしまう。 ※英日の言語的差異により出現順序が逆となるものであり、必須項目「F. 語の誤訳」には該当しない（item も them も語レベルでは正しく訳されている）。このため、カウントは「G. その他の翻訳ミス」に行う。	該当項目	EJ-08

＜EJ-09. 慣用表現とみなすべき場合／そうでない場合を適切に翻訳できていない＞

同一の表現が慣用表現として使われている場合とそうでない場合とがあります。この訳し分けが適切でない場合、ここにカウントします。

以下に、慣用表現として訳すべきでない語が慣用表現として訳されている例を挙げます。

原 文	This facilitates loading and unloading the container and also makes the container adaptable to carry differing quantities of windscreens or windscreens of differing size and curvature configuration.		
人手翻訳	これはコンテナの荷積みおよび荷降ろしを容易にし、コンテナをさまざまな 数量の フロントガラス、またはさまざまなサイズおよび曲り方の構成のフロントガラスを運ぶのに適したものとします。		
機械翻訳	…異なる 多量の フロントガラスを運ぶのに適応…		
翻訳ミス	文意を考えると指摘箇所は「異なる数量のフロントガラスを運ぶ」と訳すのが正しいが、「quantities of : 多量の」という慣用表現が優先されてしまっている。 ※quantities はここでは一般名詞として扱われるべきところ、形容詞（慣用表現 quantities of ～）と解釈されているため、品詞の誤り（D）にも該当する。	該当項目	EJ-09

一方、慣用表現として訳すべき語がそう訳されていない場合もあります。こうした場合も本項目に該当します。

原 文	[0002] A number of devices have been proposed for suspending washed laundry for drying.		
人手翻訳	【 0 0 0 2 】洗濯された洗濯物を乾燥のためにかけるための 複数の 装置が提案されている。		
機械翻訳	[0 0 0 2] 装置 の数 は、乾燥のための洗浄洗濯物を懸架する提案されている。		
翻訳ミス	A number of は「複数の、多数の」という意味の形容詞となる慣用表現だが、機械翻訳ではそう解釈せず、number を独立した名詞と解釈してしまった。 ※前例の逆のパターンだが、こちらも品詞の誤り（D）に相当する。	該当項目	EJ-09

つまり、慣用表現がそのように扱われていない場合は常に本項目に該当することとなります。なお慣用表現には、語と前置詞の組み合わせにより語または前置詞の意味が特定されるようなもの（例：upon request、rotate about）も含みます。

＜EJ-10. 前置詞に対し適切な訳語を選択できていない＞

前置詞の適訳は、述語や目的語の属性等により変化するため、対応が難しい場合があります。例えば **for** は前後の文脈に応じて「～のために」「～用の」「～という理由で」「～に向かって」など多数の訳語に訳し分ける必要があります。この基準には、前置詞が適切に訳されない翻訳ミスをカウントします。

原 文	The cylindrically curved controllable platform 128 can be rotated about its axis in either a forward or backward direction to provide material deposition in either direction.		
人手翻訳	シリンダ状に湾曲した可変台 128 は、その軸の 周囲 を前後どちらの方向にも回転して両方向に材料を被覆することができる。		
機械翻訳	一方の方角に資料被覆を提供するために、円筒状に曲がった制御可能なプラットフォーム 128 は、前方あるいは逆方向のいずれかの中のその軸芯 に関して 回転することができる。		
翻訳ミス	前置詞 about はここでは「～の周囲を」等と訳すべきである。 ※語の誤訳（F）の一種。	該当項目	EJ-12

＜EJ-11. 接頭辞、接尾辞を伴う語を適切に処理していない＞

英語特許文献には、任意の語の前後に接続して意味を追加する接頭辞・接尾辞が多用されます。主な接頭辞・接尾辞の例としては、**de-XXX**、**non-XXX**、**XXX-free**、**XXXly** などが挙げられます。

こうした接頭辞や接尾辞を伴う語が、接頭辞・接尾辞が原因で適切に翻訳されていないとみなされる場合、ここにカウントしてください。

原 文	The invention relates to a testing method for designing a semiautogenous or an autogenous grinding circuit with at least one ball mill for grinding ore.		
人手翻訳	本発明は、少なくとも 1 つのボールミルを有する 半自生 粉砕回路または自生粉砕回路を設計するための試験方法に関する。		
機械翻訳	本発明は研磨する鉱石のために少なくとも 1 個のボールミルで semiautogenous または自生粉砕回路を設計するための試験法に関する。		
翻訳ミス	文中 autogenous は「自生」と訳されているが、接頭辞 semi を伴う semiautogenous は未知語となっている。 ※未知語（C）にもカウントする。	該当項目	EJ-11

なお、意味的には把握可能であっても、例えばハイフォンが残留する等、接頭辞・接尾辞の処理が不完全な場合も、本項目の対象とします。

原 文	For example, antisense molecules can be RNAs or other molecules, including peptide nucleic acids (PNAs) or non-nucleic acid molecules such as phosphorothioate derivatives that specifically bind DNA or RNA in a base pair-dependent manner.		
人手翻訳	例えば、アンチセンス分子は、RNA、または塩基対依存様式においてDNAもしくはRNAを特異的に結合する他の分子（ペプチド核酸（PNA）または 非核酸分子 （例えば、ホスホロチオエート誘導体）を含む）であり得る。		
機械翻訳	例えば、アンチセンス分子は、リボ核酸または他の分子であるかもしれない、ベース組扶養家族方法にチオリン酸塩類縁誘導体、そんなに明確にバインド DNA または RNA などのペプチド核酸(止瀉薬)または 非-核酸分子 を含んでいる。		
翻訳ミス	機械翻訳結果は non-nucleic acid の訳語を「非-核酸分子」としており、意味は正しく把握できるがハイフォンの処理が不完全であるため本項目に該当する。 ※重要技術用語の評価における評価 B（意味は正しく把握できるが同義語の範疇を外れる語）に該当するものであり、必須項目においては語の誤訳（F）となる。	該当項目	EJ-11

<EJ-12. スラッシュを用いた並列表現を適切に処理していない>

スラッシュを用いて名詞や動詞、前置詞等を並列させる表現も多用されます。こうした表現が適切に処理されていない場合、ここにカウントします。

原 文	11. A perimeter barrier according to any preceding claim, wherein one or more additional wire ropes is/are provided, unwoven or woven relative to the intermediate post.		
人手翻訳	11. 前記間柱に対して編まれていないかまたは編まれた、1つまたはそれ以上の追加のワイヤロープが提供される、先行する請求項のいずれかに記載の周辺防壁。		
機械翻訳	11. いかなる以前の請求項に記載の周辺部バリアも、そこにおいて、中間ポストと関連して設けられているか、ほどかれるかまたは、織られる一つ以上の付加的なワイヤロープ i s / a r e 。		
翻訳ミス	動詞 is, are は基本語だが、スラッシュで連結されることで一語とみなされ、未知語扱いされてしまった。 ※未知語（C）にもカウントする。	該当項目	EJ-12

＜EJ-13. therein、wherein、whereby 等を適切に処理していない＞

英語特許文献では、タイトル部・クレーム部をはじめ各所にて、therein, therefor, wherein といった自然文ではあまり用いられない副詞が多用されます。こうした副詞を適切に表現できていない場合、本項目にカウントします。

原 文	HYDROGENATION CATALYSTS AND PROCESSES THEREWITH		
人手翻訳	水素添加触媒および これを用いた 工程		
機械翻訳	水素化触媒とプロセス それを用いた		
翻訳ミス	訳語自体は正しいが、うまく係り受けできていない。 ※係り受け誤り（E）にも該当。文として訳されているわけではない ので Z-04 には該当しない。	該当項目	EJ-13

原 文	4. Electrically conductive textile yarn (1) according to the preceding claim, wherein said core (5) has a weight which is negligible compared with the weight of said layer (6).		
人手翻訳	4. 前記コア（５）の重量が前記層（６）の重量に比べて無視し得る、先行請求項に記載の導電性繊維糸（１）。		
機械翻訳	4。一項に記載の導電性の繊維糸（１）が、前記コア（５）は層（６）を前記の重量と比較して無視できる量を有する。		
翻訳ミス	wherein の場合、この語にて接続される前後部分の関係が保たれていればどのような表現であっても OK とする。上記機械翻訳は名詞句を意図的に文として訳しており、その結果 wherein の前後部分の配置が逆転しているが、関係性は保たれているので問題ない。	該当項目	—

＜EJ-14. 代名詞 same、the same を適切に処理していない（【発明の名称】）＞

英語特許文献では、「発明の名称」部において、same や the same の代名詞用法が多用されています。この用法に適切に対応できていない場合、ここにカウントします。

原 文	Combination pouch and release sheet and method for producing same		
人手翻訳	組合せ式パウチ、剥離シート、及び その 製造方法		
機械翻訳	コンビネーションポーチ、分離型シート、および 同じ を生産する方法		
翻訳ミス	Same はここでは「これ」と訳すべきだが、対応できていない。 ※same は代名詞として取り扱うべきだが、機械翻訳では「同じ」という訳語から判断して形容詞と解釈しているため、品詞誤り（D）に該当する。	該当項目	EJ-14

<EJ-15. 小文字を含む大文字表記を適切に処理していない（【発明の名称】）>

英語特許文献の「発明の名称」は全文字大文字で表記されることが多いですが、まれに、全文字大文字表記でありながら、元素記号など一部のみ小文字が用いられるケースがあります。これが原因で入力文が適切に処理されないと思なされる場合、この基準にカウントします。

原 文	BACTERIUM COMPRISING A REGULATED rfaH NUCLEIC ACID		
人手翻訳	調整された rfaH ヌクレイン酸からなるバクテリア		
機械翻訳	BACTERIUM COMPRISING A REGULATED rfaH ヌクレイン ACID		
翻訳ミス	rfaH が小文字混じりのため大文字文とみなされず、大文字部分の大半が記号と判断された。rfaH を大文字にすれば「規制された RFAH 核酸を含む細菌」となる。 ※各語が未知語扱いされているわけではなくエンジン上の処理エラーと見なせるため、「G. その他の翻訳ミス」に 999 をセットする（添付資料 4-1 の注記参照）。	該当項目	EJ-15

<EJ-16. 引用クレームに関する表現を適切に翻訳していない>

「特許請求の範囲」において、前出クレームを引用するいくつかの定型表現があります。こうした表現に適切に対応できていない場合、ここに該当します。

原 文	10. a container according to any preceding claim , wherein a lower portion of the container includes frame openings for receiving tines of lifting apparatus.		
人手翻訳	【請求項 10】 前記コンテナの下側部分は昇降用装置の先端部を受け入れるためのフレーム開口を備える、 先行請求項のいずれか 1 項に記載の コンテナ。		
機械翻訳	10. どんな前のクレームに記載の コンテナ; そこでは、コンテナの下位部が浮き器具のタインを受けるためのフレーム開始を含んでいる。		
翻訳ミス	頻出表現であり、「前記いずれかの請求項に記載の」等と訳すのが適切。 ※any preceding claim で一語とみなし、この語の誤訳 (F) と判定。 なお、特許特有用語の誤訳 (Z-05) にも該当する。	該当項目	EJ-16

以 上

添付資料7-7

[illegible]

添付資料7-7. 品質評価結果データ データサンプル

添付資料7-7

No.	文番号	文庫番号	文庫記号	対応番号	項目	IPC	原文	語数 文字数	人手翻訳	機械翻訳結果	構文 翻訳ID	メモ欄	内容伝達①	内容伝達②	内容伝達③	平均値	短文	中文	長文	重要技術用語①	用語① 評価 A-D	A	B	C	D	重要技術用語②	用語② 評価 A-D	A	B	C	D	重要技術用語③	用語③ 評価 A-D	A	B	C	D	重要技術用語④	用語④ 評価 A-D	A	B	C	D	重要技術用語⑤	用語⑤ 評価 A-D	A	B	C	D	重要技術用語⑥	用語⑥ 評価 A-D	A	B	C	D	重要技術用語⑦	用語⑦ 評価 A-D	A	B	C	D	重要技術用語⑧	用語⑧ 評価 A-D	A	B	C	D	重要技術用語⑨	用語⑨ 評価 A-D	A	B	C	D	重要技術用語⑩	用語⑩ 評価 A-D	A	B	C	D	重要技術用語⑪	用語⑪ 評価 A-D	A	B	C	D	重要技術用語⑫	用語⑫ 評価 A-D	A	B	C	D	重要技術用語⑬	用語⑬ 評価 A-D	A	B	C	D	重要技術用語⑭	用語⑭ 評価 A-D	A	B	C	D	重要技術用語⑮	用語⑮ 評価 A-D	A	B	C	D	重要技術用語⑯	用語⑯ 評価 A-D	A	B	C	D	重要技術用語⑰	用語⑰ 評価 A-D	A	B	C	D	重要技術用語⑱	用語⑱ 評価 A-D	A	B	C	D	重要技術用語⑲	用語⑲ 評価 A-D	A	B	C	D	重要技術用語⑳	用語㉑ 評価 A-D	A	B	C	D	重要技術用語㉒	用語㉓ 評価 A-D	A	B	C	D	重要技術用語㉔	用語㉕ 評価 A-D	A	B	C	D	重要技術用語㉖	用語㉗ 評価 A-D	A	B	C	D	重要技術用語㉘	用語㉙ 評価 A-D	A	B	C	D	重要技術用語㉚	用語㉛ 評価 A-D	A	B	C	D	重要技術用語㉜	用語㉝ 評価 A-D	A	B	C	D	重要技術用語㉞	用語㉟ 評価 A-D	A	B	C	D	重要技術用語㊱	用語㊲ 評価 A-D	A	B	C	D	重要技術用語㊳	用語㊴ 評価 A-D	A	B	C	D	重要技術用語㊵	用語㊶ 評価 A-D	A	B	C	D	重要技術用語㊷	用語㊸ 評価 A-D	A	B	C	D	重要技術用語㊹	用語㊺ 評価 A-D	A	B	C	D	重要技術用語㊻	用語㊼ 評価 A-D	A	B	C	D	重要技術用語㊽	用語㊾ 評価 A-D	A	B	C	D	重要技術用語㊿	用語㋀ 評価 A-D	A	B	C	D	重要技術用語㋁	用語㋂ 評価 A-D	A	B	C	D	重要技術用語㋃	用語㋄ 評価 A-D	A	B	C	D	重要技術用語㋅	用語㋆ 評価 A-D	A	B	C	D	重要技術用語㋇	用語㋈ 評価 A-D	A	B	C	D	重要技術用語㋉	用語㋊ 評価 A-D	A	B	C	D	重要技術用語㋋	用語㋌ 評価 A-D	A	B	C	D	重要技術用語㋍	用語㋎ 評価 A-D	A	B	C	D	重要技術用語㋏	用語㋐ 評価 A-D	A	B	C	D	重要技術用語㋑	用語㋒ 評価 A-D	A	B	C	D	重要技術用語㋓	用語㋔ 評価 A-D	A	B	C	D	重要技術用語㋕	用語㋖ 評価 A-D	A	B	C	D	重要技術用語㋗	用語㋘ 評価 A-D	A	B	C	D	重要技術用語㋙	用語㋚ 評価 A-D	A	B	C	D	重要技術用語㋛	用語㋜ 評価 A-D	A	B	C	D	重要技術用語㋝	用語㋞ 評価 A-D	A	B	C	D	重要技術用語㋟	用語㋠ 評価 A-D	A	B	C	D	重要技術用語㋡	用語㋢ 評価 A-D	A	B	C	D	重要技術用語㋣	用語㋤ 評価 A-D	A	B	C	D	重要技術用語㋥	用語㋦ 評価 A-D	A	B	C	D	重要技術用語㋧	用語㋨ 評価 A-D	A	B	C	D	重要技術用語㋩	用語㋪ 評価 A-D	A	B	C	D	重要技術用語㋫	用語㋬ 評価 A-D	A	B	C	D	重要技術用語㋭	用語㋮ 評価 A-D	A	B	C	D	重要技術用語㋯	用語㋰ 評価 A-D	A	B	C	D	重要技術用語㋱	用語㋲ 評価 A-D	A	B	C	D	重要技術用語㋳	用語㋴ 評価 A-D	A	B	C	D	重要技術用語㋵	用語㋶ 評価 A-D	A	B	C	D	重要技術用語㋷	用語㋸ 評価 A-D	A	B	C	D	重要技術用語㋹	用語㋺ 評価 A-D	A	B	C	D	重要技術用語㋻	用語㋼ 評価 A-D	A	B	C	D	重要技術用語㋽	用語㋾ 評価 A-D	A	B	C	D	重要技術用語㋿	用語㊀ 評価 A-D	A	B	C	D	重要技術用語㊁	用語㊂ 評価 A-D	A	B	C	D	重要技術用語㊃	用語㊄ 評価 A-D	A	B	C	D	重要技術用語㊅	用語㊆ 評価 A-D	A	B	C	D	重要技術用語㊇	用語㊈ 評価 A-D	A	B	C	D	重要技術用語㊉	用語㊊ 評価 A-D	A	B	C	D	重要技術用語㊋	用語㊌ 評価 A-D	A	B	C	D	重要技術用語㊍	用語㊎ 評価 A-D	A	B	C	D	重要技術用語㊏	用語㊐ 評価 A-D	A	B	C	D	重要技術用語㊑	用語㊒ 評価 A-D	A	B	C	D	重要技術用語㊓	用語㊔ 評価 A-D	A	B	C	D	重要技術用語㊕	用語㊖ 評価 A-D	A	B	C	D	重要技術用語㊗	用語㊘ 評価 A-D	A	B	C	D	重要技術用語㊙	用語㊚ 評価 A-D	A	B	C	D	重要技術用語㊛	用語㊜ 評価 A-D	A	B	C	D	重要技術用語㊝	用語㊞ 評価 A-D	A	B	C	D	重要技術用語㊟	用語㊠ 評価 A-D	A	B	C	D	重要技術用語㊡	用語㊢ 評価 A-D	A	B	C	D	重要技術用語㊣	用語㊤ 評価 A-D	A	B	C	D	重要技術用語㊥	用語㊦ 評価 A-D	A	B	C	D	重要技術用語㊧	用語㊨ 評価 A-D	A	B	C	D	重要技術用語㊩	用語㊪ 評価 A-D	A	B	C	D	重要技術用語㊫	用語㊬ 評価 A-D	A	B	C	D	重要技術用語㊭	用語㊮ 評価 A-D	A	B	C	D	重要技術用語㊯	用語㊰ 評価 A-D	A	B	C	D	重要技術用語㊱	用語㊲ 評価 A-D	A	B	C	D	重要技術用語㊳	用語㊴ 評価 A-D	A	B	C	D	重要技術用語㊵	用語㊶ 評価 A-D	A	B	C	D	重要技術用語㊷	用語㊸ 評価 A-D	A	B	C	D	重要技術用語㊹	用語㊺ 評価 A-D	A	B	C	D	重要技術用語㊻	用語㊼ 評価 A-D	A	B	C	D	重要技術用語㊽	用語㊾ 評価 A-D	A	B	C	D	重要技術用語㊿	用語㋀ 評価 A-D	A	B	C	D	重要技術用語㋁	用語㋂ 評価 A-D	A	B	C	D	重要技術用語㋃	用語㋄ 評価 A-D	A	B	C	D	重要技術用語㋅	用語㋆ 評価 A-D	A	B	C	D	重要技術用語㋇	用語㋈ 評価 A-D	A	B	C	D	重要技術用語㋉	用語㋊ 評価 A-D	A	B	C	D	重要技術用語㋋	用語㋌ 評価 A-D	A	B	C	D	重要技術用語㋍	用語㋎ 評価 A-D	A	B	C	D	重要技術用語㋏	用語㋐ 評価 A-D	A	B	C	D	重要技術用語㋑	用語㋒ 評価 A-D	A	B	C	D	重要技術用語㋓	用語㋔ 評価 A-D	A	B	C	D	重要技術用語㋕	用語㋖ 評価 A-D	A	B	C	D	重要技術用語㋗	用語㋘ 評価 A-D	A	B	C	D	重要技術用語㋙	用語㋚ 評価 A-D	A	B	C	D	重要技術用語㋛	用語㋜ 評価 A-D	A	B	C	D	重要技術用語㋝	用語㋞ 評価 A-D	A	B	C	D	重要技術用語㋟	用語㋠ 評価 A-D	A	B	C	D	重要技術用語㋡	用語㋢ 評価 A-D	A	B	C	D	重要技術用語㋣	用語㋤ 評価 A-D	A	B	C	D	重要技術用語㋥	用語㋦ 評価 A-D	A	B	C	D	重要技術用語㋧	用語㋨ 評価 A-D	A	B	C	D	重要技術用語㋩	用語㋪ 評価 A-D	A	B	C	D	重要技術用語㋫	用語㋬ 評価 A-D	A	B	C	D	重要技術用語㋭	用語㋮ 評価 A-D	A	B	C	D	重要技術用語㋯	用語㋰ 評価 A-D	A	B	C	D	重要技術用語㋱	用語㋲ 評価 A-D	A	B	C	D	重要技術用語㋳	用語㋴ 評価 A-D	A	B	C	D	重要技術用語㋵	用語㋶ 評価 A-D	A	B	C	D	重要技術用語㋷	用語㋸ 評価 A-D	A	B	C	D	重要技術用語㋹	用語㋺ 評価 A-D	A	B	C	D	重要技術用語㋻	用語㋼ 評価 A-D	A	B	C	D	重要技術用語㋽	用語㋾ 評価 A-D	A	B	C	D	重要技術用語㋿	用語㊀ 評価 A-D	A	B	C	D	重要技術用語㊁	用語㊂ 評価 A-D	A	B	C	D	重要技術用語㊃	用語㊄ 評価 A-D	A	B	C	D	重要技術用語㊅	用語㊆ 評価 A-D	A	B	C	D	重要技術用語㊇	用語㊈ 評価 A-D	A	B	C	D	重要技術用語㊉	用語㊊ 評価 A-D	A	B	C	D	重要技術用語㊋	用語㊌ 評価 A-D	A	B	C	D	重要技術用語㊍	用語㊎ 評価 A-D	A	B	C	D	重要技術用語㊏	用語㊐ 評価 A-D	A	B	C	D	重要技術用語㊑	用語㊒ 評価 A-D	A	B	C	D	重要技術用語㊓	用語㊔ 評価 A-D	A	B	C	D	重要技術用語㊕	用語㊖ 評価 A-D	A	B	C	D	重要技術用語㊗	用語㊘ 評価 A-D	A	B	C	D	重要技術用語㊙	用語㊚ 評価 A-D	A	B	C	D	重要技術用語㊛	用語㊜ 評価 A-D	A	B	C	D	重要技術用語㊝	用語㊞ 評価 A-D	A	B	C	D	重要技術用語㊟	用語㊠ 評価 A-D	A	B	C	D	重要技術用語㊡	用語㊢ 評価 A-D	A	B	C	D	重要技術用語㊣	用語㊤ 評価 A-D	A	B	C	D	重要技術用語㊥	用語㊦ 評価 A-D	A	B	C	D	重要技術用語㊧	用語㊨ 評価 A-D	A	B	C	D	重要技術用語㊩	用語㊪ 評価 A-D	A	B	C	D	重要技術用語㊫	用語㊬ 評価 A-D	A	B	C	D	重要技術用語㊭	用語㊮ 評価 A-D	A	B	C	D	重要技術用語㊯	用語㊰ 評価 A-D	A	B	C	D	重要技術用語㊱	用語㊲ 評価 A-D	A	B	C	D	重要技術用語㊳	用語㊴ 評価 A-D	A	B	C	D	重要技術用語㊵	用語㊶ 評価 A-D	A	B	C	D	重要技術用語㊷	用語㊸ 評価 A-D	A	B	C	D	重要技術用語㊹	用語㊺ 評価 A-D	A	B	C	D	重要技術用語㊻	用語㊼ 評価 A-D	A	B	C	D	重要技術用語㊽	用語㊾ 評価 A-D	A	B	C	D	重要技術用語㊿	用語㋀ 評価 A-D	A	B	C	D	重要技術用語㋁	用語㋂ 評価 A-D	A	B	C	D	重要技術用語㋃	用語㋄ 評価 A-D	A	B	C	D	重要技術用語㋅	用語㋆ 評価 A-D	A	B	C	D	重要技術用語㋇	用語㋈ 評価 A-D	A	B	C	D	重要技術用語㋉	用語㋊ 評価 A-D	A	B	C	D	重要技術用語㋋	用語㋌ 評価 A-D	A	B	C	D	重要技術用語㋍	用語㋎ 評価 A-D	A	B	C	D	重要技術用語㋏	用語㋐ 評価 A-D	A	B	C	D	重要技術用語㋑	用語㋒ 評価 A-D	A	B	C	D	重要技術用語㋓	用語㋔ 評価 A-D	A	B	C	D	重要技術用語㋕	用語㋖ 評価 A-D	A	B	C	D	重要技術用語㋗	用語㋘ 評価 A-D	A	B	C	D	重要技術用語㋙	用語㋚ 評価 A-D	A	B	C	D	重要技術用語㋛	用語㋜ 評価 A-D	A	B	C	D	重要技術用語㋝	用語㋞ 評価 A-D	A	B	C	D	重要技術用語㋟	用語㋠ 評価 A-D	A	B	C	D	重要技術用語㋡	用語㋢ 評価 A-D	A	B	C	D	重要技術用語㋣	用語㋤ 評価 A-D	A	B	C	D	重要技術用語㋥	用語㋦ 評価 A-D	A	B	C	D	重要技術用語㋧	用語㋨ 評価 A-D	A	B	C	D	重要技術用語㋩	用語㋪ 評価 A-D	A	B	C	D	重要技術用語㋫	用語㋬ 評価 A-D	A	B	C	D	重要技術用語㋭	用語㋮ 評価 A-D	A	B	C	D	重要技術用語㋯	用語㋰ 評価 A-D	A	B	C	D	重要技術用語㋱	用語㋲ 評価 A-D	A	B	C	D	重要技術用語㋳	用語㋴ 評価 A-D	A	B	C	D	重要技術用語㋵	用語㋶ 評価 A-D	A	B	C	D	重要技術用語㋷	用語㋸ 評価 A-D	A	B	C	D	重要技術用語㋹	用語㋺ 評価 A-D	A	B	C	D	重要技術用語㋻	用語㋼ 評価 A-D	A	B	C	D	重要技術用語㋽	用語㋾ 評価 A-D	A	B	C	D	重要技術用語㋿	用語㊀ 評価 A-D	A	B	C	D	重要技術用語㊁	用語㊂ 評価 A-D	A	B	C	D	重要技術用語㊃	用語㊄ 評価 A-D	A	B	C	D	重要技術用語㊅	用語㊆ 評価 A-D	A	B	C	D	重要技術用語㊇	用語㊈ 評価 A-D	A	B	C	D	重要技術用語㊉	用語㊊ 評価 A-D	A	B	C	D	重要技術用語㊋	用語㊌ 評価 A-D	A	B	C	D	重要技術用語㊍	用語㊎ 評価 A-D	A	B	C	D	重要技術用語㊏	用語㊐ 評価 A-D	A	B	C	D	重要技術用語㊑	用語㊒ 評価 A-D	A	B	C	D	重要技術用語㊓	用語㊔ 評価 A-D	A	B	C	D	重要技術用語㊕	用語㊖ 評価 A-D	A	B	C	D	重要技術用語㊗	用語㊘ 評価 A-D	A	B	C	D	重要技術用語㊙	用語㊚ 評価 A-D	A	B	C	D	重要技術用語㊛	用語㊜ 評価 A-D	A	B	C	D	重要技術用語㊝	用語㊞ 評価 A-D	A	B	C	D	重要技術用語㊟	用語㊠ 評価 A-D	A	B	C	D	重要技術用語㊡	用語㊢ 評価 A-D	A	B	C	D	重要技術用語㊣	用語㊤ 評価 A-D	A	B	C	D	重要技術用語㊥	用語㊦ 評価 A-D	A	B	C	D	重要技術用語㊧	用語㊨ 評価 A-D	A	B	C	D	重要技術用語㊩	用語㊪ 評価 A-D	A	B	C	D	重要技術用語㊫	用語㊬ 評価 A-D	A	B	C	D	重要技術用語㊭	用語㊮ 評価 A-D	A	B	C	D	重要技術用語㊯	用語㊰ 評価 A-D	A	B	C	D	重要技術用語㊱	用語㊲ 評価 A-D	A	B	C	D	重要技術用語㊳	用語㊴ 評価 A-D	A	B	C	D	重要技術用語㊵	用語㊶ 評価 A-D	A	B	C	D	重要技術用語㊷	用語㊸ 評価 A-D	A	B	C	D	重要技術用語㊹	用語㊺ 評価 A-D	A	B	C	D	重要技術用語㊻	用語㊼ 評価 A-D	A	B	C	D	重要技術用語㊽	用語㊾ 評価 A-D	A	B	C	D	重要技術用語㊿	用語㋀ 評価 A-D	A	B	C	D	重要技術用語㋁	用語㋂ 評価 A-D	A	B	C	D	重要技術用語㋃	用語㋄ 評価 A-D	A	B	C	D	重要技術用語㋅	用語㋆ 評価 A-D	A	B	C	D	重要技術用語㋇	用語㋈ 評価 A-D	A	B	C	D	重要技術用語㋉	用語㋊ 評価 A-D	A	B	C	D	重要技術用語㋋	用語㋌ 評価 A-D	A	B	C	D	重要技術用語㋍	用語㋎ 評価 A-D	A	B	C	D	重要技術用語㋏	用語㋐ 評価 A-D	A	B	C	D	重要技術用語㋑	用語㋒ 評価 A-D	A	B	C	D	重要技術用語㋓	用語㋔ 評価 A-D	A	B	C	D	重要技術用語㋕	用語㋖ 評価 A-D	A	B	C	D	重要技術用語㋗	用語㋘ 評価 A-D	A	B	C	D	重要技術用語㋙	用語㋚ 評価 A-D	A	B	C	D	重要技術用語㋛	用語㋜ 評価 A-D	A	B	C	D	重要技術用語㋝	用語㋞ 評価 A-D	A	B	C	D	重要技術用語㋟	用語㋠ 評価 A-D	A	B	C	D	重要技術用語㋡	用語㋢ 評価 A-D	A	B	C	D	重要技術用語㋣	用語㋤ 評価 A-D	A	B	C	D	重要技術用語㋥	用語㋦ 評価 A-D	A	B	C	D	重要技術用語㋧	用語㋨ 評価 A-D	A	B	C	D	重要技術用語㋩	用語㋪ 評価 A-D	A	B	C	D	重要技術用語㋫	用語㋬ 評価 A-D	A	B	C	D	重要技術用語㋭	用語㋮ 評価 A-D	A	B	C	D	重要技術用語㋯	用語㋰ 評価 A-D	A	B	C	D	重要技術用語㋱	用語㋲ 評価 A-D	A	B	C	D	重要技術用語㋳	用語㋴ 評価 A-D	A	B	C	D	重要技術用語㋵	用語㋶ 評価 A-D	A	B	C	D	重要技術用語㋷	用語㋸ 評価 A-D	A	B	C	D	重要技術用語㋹	用語㋺ 評価 A-D	A	B	C	D	重要技術用語㋻	用語㋼ 評価 A-D	A	B	C	D	重要技術用語㋽	用語㋾ 評価 A-D	A	B	C	D	重要技術用語㋿	用語㊀ 評価 A-D	A	B	C	D	重要技術用語㊁	用語㊂ 評価 A-D	A	B	C	D	重要技術用語㊃	用語㊄ 評価 A-D	A	B	C
-----	-----	------	------	------	----	-----	----	-----------	------	--------	------------	-----	-------	-------	-------	-----	----	----	----	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---	---	---------	------------------	---	---	---

自動評価の粗ぶるいへの利用適否実験

1. 自動評価の実行

本調査の品質評価実践に使用した 5 つの機械翻訳結果(RBMT1～SMT5)および人工的に作成した質の悪い機械翻訳結果(MT0)に対して自動評価(BLEU、NIST、RIBES)を行った。対象文数は 199 文である。ここで MT0 は以下の翻訳結果である。

MT0 : google の英中翻訳+J 北京の中日翻訳

前処理として、リファレンス（人手翻訳）と、6 種類の機械翻訳結果に対して、以下の処理を施した。

- ・半角全角変換

NTCIR-10 でオーガナイザが用いた `h2z-utf8-without-space.pl`¹ を用いて半角文字を全角文字に変換した。これは半角と全角の違いによる評価値の低下を除くためである。

- ・形態素分割

ChaSen を用いて形態素分割を行った。

2. BLEU、NIST、RIBES スコアの計算

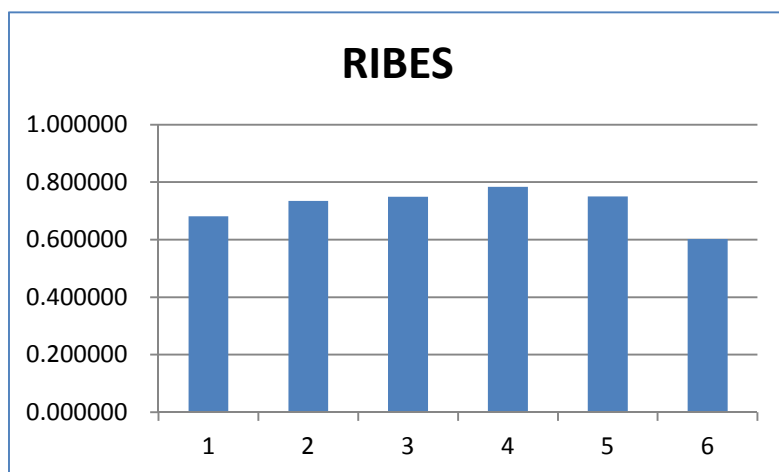
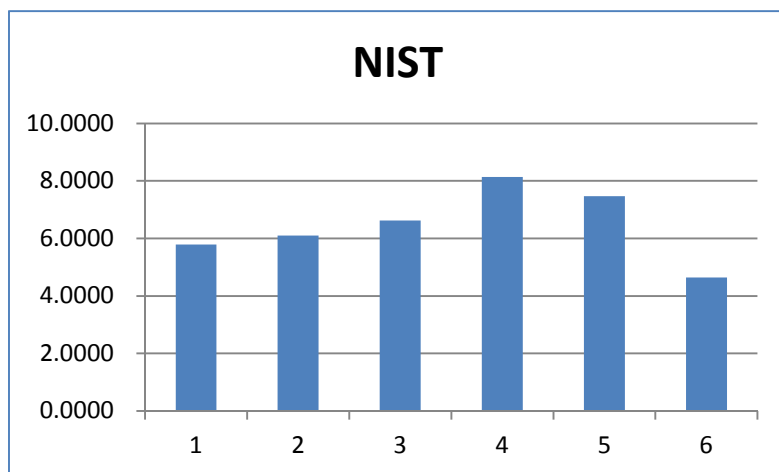
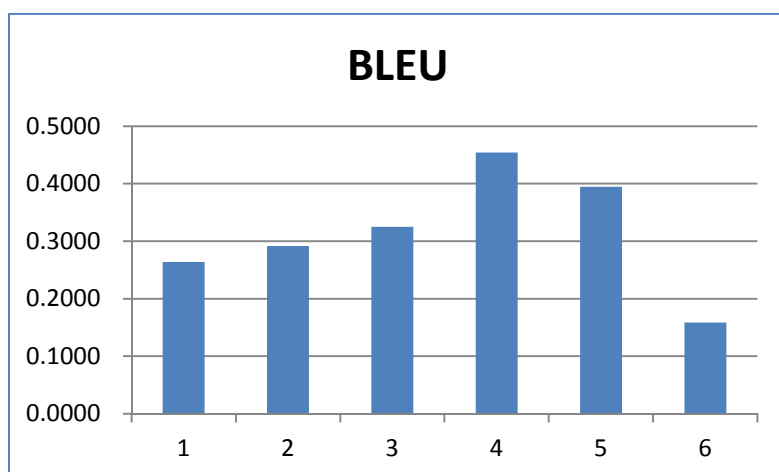
NIST が開発した `mteval-v13.pl`² を用いて BLEU と NIST のスコアを求めた。結果を下表および下図に示す。ここに示すスコアは system レベルの Cumulative 4-gram score である。また、RIBES-1.02.4³を用いて RIBES を計算した。結果をあわせて下表および下図に示す。

システム名	BLEU	NIST	RIBES
RBMT1	0.2640	5.7864	0.681150
RBMT2	0.2917	6.1002	0.734709
RBMT3	0.3253	6.6208	0.749215
SMT4	0.4543	8.1341	0.783155
SMT5	0.3947	7.4643	0.749765
MT0	0.1585	4.6428	0.602517

¹ http://ntcir.nii.ac.jp/PatentMT-2/?action=common_download_main&upload_id=322

² <ftp://jaguar.ncsl.nist.gov/mt/resources/mteval-v13.pl>

³ <http://www.kecl.ntt.co.jp/icl/lirg/ribes/index-j.html>



※各グラフとも、1～5：RBMT1～SMT5、6：MT0。

機械翻訳エンジンにおける一般的なカスタマイズ機能例

1. 用語知識の登録

機械翻訳エンジンの知識に収録されていない用語の情報を、新たな知識として追加登録する手法であり、カスタマイズ機能の最も基本的な機能といえる。

原 文	Plunger 95 is mounted pivotally to slide block 92 at the proximal end and slidably disposed within plunger housing at a distal end.
機械翻訳	向こうみずの相場師 95 は、近位末端で遮断 92 を滑らせるように重要に取り付けられて、向こうみずの相場師ハウジングの中に slidably に遠位末端に配置される。

上例の機械翻訳結果は、原文の内容を十分に伝えているとはいえず、「内容の伝達レベル」の評価値では 1 に相当する。このような翻訳品質の低い機械翻訳結果を、カスタマイズ機能のひとつである用語知識の登録によって改善する具体例を示す。

まず、**slidably** は辞書に収録されていないために、原文表記のまま訳文に出力されている。これを副詞としてユーザ辞書に追加登録し、「スライド可能に」という訳語を付与する。また、「向こうみずの相場師」と訳されている **plunger** についても、同様に「プランジャ」という訳語で名詞として追加登録する。同じく「重要に」と訳されている **pivotally** も、特許文献では「枢動可能に」の意味で用いられることが圧倒的に多いため、こちらの訳語で辞書登録する。

また、原文の **slide block** は「スライドブロック」という技術用語を表す英語表現だが、これがエンジンの辞書に収録されていないために、**slide** が動詞と解釈され、**slide block 92** が「遮断 92 を滑らせる」と誤訳されている。これは、この語 (**slide block**) をユーザ辞書に複合名詞としてこの単位で登録することで、以後上記のような品詞誤りを回避できる。

以上のような用語知識の登録により、機械翻訳結果は以下のように改善されることが予想される。

機械翻訳 (カスタマイズ 後)	プランジャ 95 は、近位末端でスライドブロック 92 に枢動可能に取り付けられて、プランジャハウジングの中にスライド可能に遠位末端に配置される。
--------------------	---

上記翻訳結果の内容伝達スコアは 4 に相当し、大幅な改善がみられる。このような用語知識の登録は、特に RBMT 方式のエンジンにおいて大きな効果を示す。

2. その他の知識の登録

機械翻訳エンジンによっては、用語単位の知識ではなく、句単位の翻訳知識、文単位の翻訳知識を追加するカスタマイズ機能を持つものもある。

これらのカスタマイズ機能は、それぞれ異なるレベルで効果を表す。句単位の翻訳知識、たとえば動詞句（動詞+目的語）としての翻訳知識は、機械翻訳時にその動詞句としての解釈が得られる場合において効果を表す。従って、登録した翻訳知識が効果を発揮するためには、翻訳処理時に動詞句としての解釈が得られる必要がある。このため、句単位の登録は、SMT 方式や EBMT 方式よりも、文法規則に基づいた翻訳処理を行う方式である RBMT 方式において明確な効果を得られる可能性が高い。

一方、文単位の翻訳知識は、EBMT 方式にとっては直接的な知識であり、SMT 方式にとっては学習データの一部となりうる。特許文献において頻出する文型を翻訳知識としてそのままエンジンに登録するだけで、その文型に合致する入力文を、適切な文型で翻訳することができる。なお、RBMT 方式を基本エンジンとした機械翻訳ソフトウェアには、EBMT 方式が組み込まれているものもある。

3. 登録すべき知識の抽出

ここで示した各種翻訳知識、特に用語レベルの翻訳知識は、利用者が翻訳結果を参照して抽出する方法が一般的である。翻訳精度が十分でない訳文を調べて、追加登録が必要な知識を探し出す必要がある。

ただし、高機能な機械翻訳ソフトウェアの中には、このような知識を、半自動的に抽出する機能を有するものもある。入力原文から機械翻訳の辞書に収録されていない用語を取り出して利用者に示す機能は、その中でも基本的なものである。

また、参照すべき訳文が存在する場合に限るが、こうした用語に対する訳語を推定して提示する機能も存在する。

これらの機能を活用することにより、カスタマイズの効率を向上させることができる。

以 上

NTCIR 特許翻訳タスクにおける言語横断特許検索による Extrinsic 評価実験に関する調査

NTCIR-7 特許機械翻訳タスク[1]および NTCIR-8 特許機械翻訳タスク[2]での、言語横断特許検索による Extrinsic 評価実験について説明する。両タスクとも、ほぼ同様の手順・仕様で評価が行われたが、本節では特に、より詳細な評価・分析結果が報告されている NTCIR-7 特許機械翻訳タスクについて報告する。

1. 評価手順

評価手順を図 1 に示す。Extrinsic 評価においては、評価対象 MT システムによって英日翻訳された結果の和訳文を用いて日本語特許文検索を行うという言語横断特許検索タスクの検索精度によって、間接的に MT システムの翻訳精度を評価する。特許検索精度評価の際には、NTCIR-5 の特許検索タスク[3]において構築したテストコレクションを使用した。全 1,189 個の検索課題(請求項)のうち、日本語における特許検索精度が中程度の 124 検索課題を用いた。また、特許検索において用いる検索システムを参加システム全体で統一し、MT システムそのものの性能差が特許検索に及ぼす影響を比較するために、タスクオーガナイザが統一検索システムを用意した。具体的には、Okapi BM25[4]を用い、各検索課題について、スコア上位のうち IPC 分類の整合性を満たす 1,000 件を評価対象とした。評価尺度としては、Mean Average Precision (MAP)、および、上位 N 位までの文書に対する再現率(Recall@N, N=100, 200, 500, 1000)を用いた。

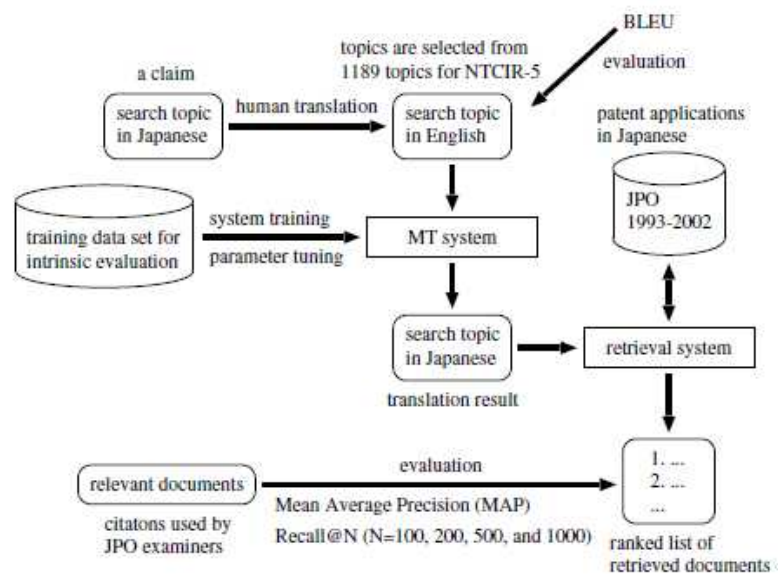


図 1 : NTCIR-7 特許機械翻訳タスクにおける
言語横断特許検索による Extrinsic 評価の手順 (文献[1]より引用)

2. 評価結果

タスク参加 12 チームに Moses[5]を加えた計 13 システムの評価結果を表 1 に示す. また, 翻訳精度と検索性能の相関を評価するために, 検索において用いた MT 訳そのものに対する Extrinsic の BLEU と特許検索精度の相関を図 2 に示す. この結果において, BLEU と MAP との間には高い相関がみられ, 相関係数は 0.936 となった[6].

さらに, 任意の 2 システムの組み合わせに対して, 検索において用いた MT 訳の BLEU の差と特許検索性能(MAP)の差の統計的有意差との間の相関をプロットした結果を図 3 に示す. この結果においては, MAP の統計的有意差が確実に 1%となるためには, BLEU において 10 ポイント以上の差が必要であることが示された. ただし, ここでの絶対値そのものは評価に用いるデータセットに依存するので, さらなる評価をする必要がある, とのことである.

表 1 : 英日方向 Intrinsic/Extrinsic 評価の結果 (文献[1]より引用)

Group	Method	Intrinsic				Extrinsic	
		BLEU	Human	Adequacy	Fluency	BLEU	MAP
Moses *	SMT	30.58	3.30	2.90	3.69	20.70	0.3140
HCRL	SMT	29.97	—	—	—	21.10	0.3536
NICT-ATR	SMT	29.15	2.89	2.59	3.20	19.40	0.3494
NTT	SMT	28.07	3.14	2.74	3.54	18.69	0.3456
NAIST-NTT	SMT	27.19	—	—	—	20.46	0.3248
KLE	SMT	26.93	—	—	—	19.07	0.2925
tori	SMT	25.33	—	—	—	17.54	0.3187
(MIBEL)	SMT	23.72	—	—	—	18.67	0.2873
HIT2	SMT	22.84	—	—	—	17.71	0.2777
(Kyoto-U)	EBMT	22.65	2.48	2.42	2.54	13.75	0.2817
(tsbmt)	RBMT	17.46	3.60	3.53	3.67	12.39	0.2264
FDU-MCandWI	SMT	10.52	—	—	—	11.10	0.2562
TH	SMT	2.23	—	—	—	1.39	0.1000
Mono	—	—	—	—	—	—	0.4797

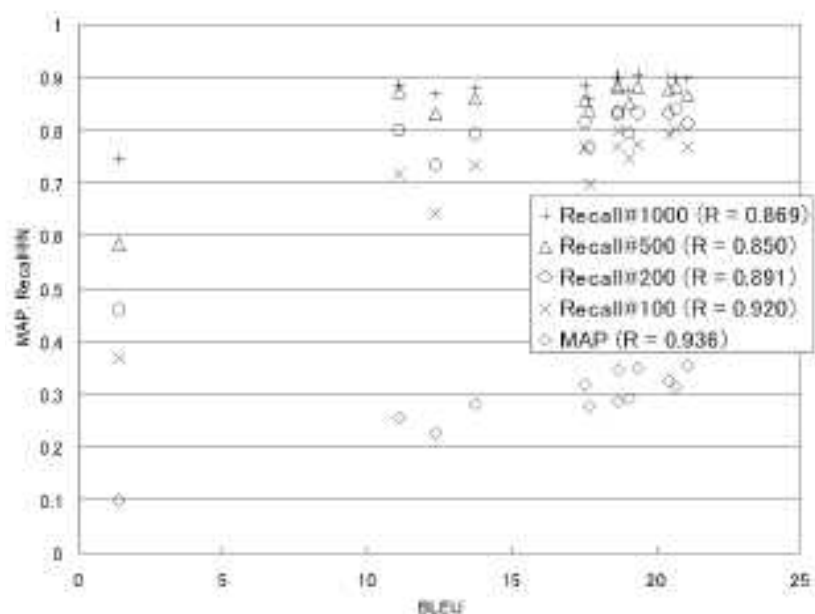


図 2：英日方向 Extrinsic 評価における BLEU と言語横断特許検索性能との相関（文献[1]より引用）

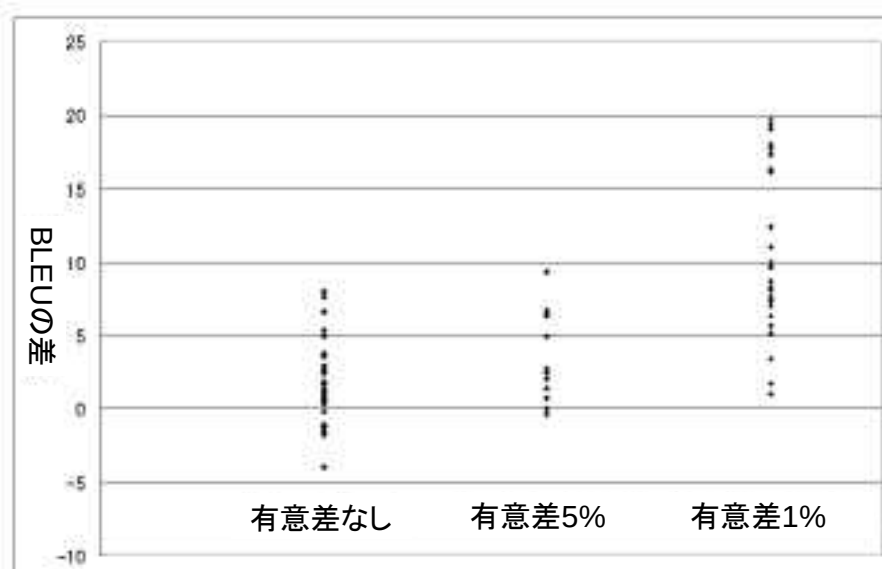


図 3：英日方向 Extrinsic 評価における BLEU の差と言語横断特許検索性能(MAP)の差の統計的有意差との相関（文献[1]より引用）

参考文献

- [1] A. Fujii, M. Utiyama, M. Yamamoto, and T. Utsuro. Overview of the Patent Translation Task at the NTCIR-7 Workshop. In *Proc. 7th NTCIR Workshop Meeting*, pp. 389-400, 2008.
- [2] A. Fujii, M. Utiyama, M. Yamamoto, T. Utsuro, H. Echizen-ya, T. Ehara, and S. Shimohata. Overview of the Patent Translation Task at the NTCIR-8 Workshop. In *Proc. 8th NTCIR Workshop Meeting*, pp. 371-376, 2010.
- [3] A. Fujii, M. Iwayama, and N. Kando. Test Collections for Patent Retrieval and Patent Classification in the 5th NTCIR Workshop, In *Proc. 5th LREC*, pp. 671-674, 2006.
- [4] S. E. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford. Okapi at TREC-3. In *Proc. 3rd TREC*, 1994.
- [5] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proc. 45th ACL, Companion Volume*, pp. 177-180, 2007.
- [6] 藤井敦, 内山将夫, 山本幹雄, 宇津呂武仁. 大規模な共通基盤による機械翻訳システムの比較評価: NTCIR 特許翻訳タスク最新事情, 言語処理学会第15回年次大会論文集, pp. 204-207, 2009