

平成 29 年度
「中国特許文献の解析及びデータ作成事業」

報 告 書

平成 30 年 3 月 28 日

一般財団法人 日本特許情報機構

目次

1. 中国文献の機械翻訳文の翻訳品質評価	3
1.1 本評価の概要	3
1.1.1 機械的手法・書誌情報による人手品質評価対象分野の選定（一次評価）	3
1.1.2 人手による品質評価の実施（二次評価）	4
1.1.3 人手による新旧機械翻訳結果の比較評価の実施（新旧 MT 比較評価）	5
1.2 一次評価の実施	6
1.2.1 一次評価対象データ	6
1.2.2 中 AU の特定	6
1.2.3 BLEU 及び RIBES による自動評価結果	7
1.2.4 補足的な評価観点の反映	14
1.2.5 二次評価対象中 AU の決定	21
1.3 二次評価の実施	23
1.3.1 評価対象文献の選定	23
1.3.2 評価対象文の選定	24
1.3.3 重要技術用語の選定	28
1.3.4 評価基準	29
1.3.5 二次評価結果（評価対象文単位）	31
1.3.6 二次評価結果の集計（中 AU 単位）	33
1.3.7 二次評価結果の分析	36
1.4 和文抄録作成対象案件の選定	57
1.4.1 テーマコード単位の選定	57
1.4.2 和文抄録作成可能案件の特定	57
1.4.3 和文抄録作成対象案件 8 万件の選定	58
1.5 新旧の機械翻訳文の比較評価の実施	60
1.5.1 新旧機械翻訳文の比較評価の概要	60
1.5.2 評価対象文献の選定	60
1.5.3 評価対象文候補の選定	61
1.5.4 評価対象文予備候補の準備	61
1.5.5 新規機械翻訳文の取得	62
1.5.6 評価対象 500 文の選定	63
1.5.7 重要技術用語の選定	64
1.5.8 評価基準	64
1.5.9 新旧機械翻訳文の変化率の分析	65

1.5.10 新旧機械翻訳文の比較評価結果	67
1.5.11 比較評価結果の分析①「特許庁自身によるチューニング」	72
1.5.12 比較評価結果の分析②「前年度事業辞書データ適用による効果」	75
1.5.12 まとめ	88
2. 特定テーマコードの中国特許公報の和文抄録の作成	89
2.1 和文抄録の作成対象	89
2.2 和文抄録の作成結果	89
2.3 特殊案件	89
3. 国際調査報告で引用された文献の全文翻訳文の作成	90
3.1 全文翻訳文の作成結果	90
3.2 原文が4万文字以上の文献	90
4. 対訳コーパスの作成	91
4.1 対訳コーパスの単位	91
4.2 対訳コーパスの作成対象	93
4.3 対訳コーパスの作成件数	94
5. 辞書データの作成	100
5.1 辞書データの作成手順	100
5.2 辞書データの作成件数	101
5.2.1 辞書データ作成件数	101
5.1.2 データフォーマット	102
6. テーマコード情報の作成	104
6.1 「テーマコード情報リスト」更新内容	104

1. 中国文献の機械翻訳文の翻訳品質評価

1.1 本評価の概要

平成 29 年度「中国特許文献の解析及びデータ作成事業」の一環として、「中韓文献翻訳・検索システム」を通じて提供している中国特許の機械翻訳データについて、二段階の翻訳品質評価を実施した。本評価の目的は、「中韓文献翻訳・検索システム」において相対的に機械翻訳精度の低い技術分野を特定することであり、上記事業は、この評価結果に基づき当該分野の中国和文抄録を作成することで、当該分野の言語資源を充実させ、そこから得られる中日対訳コーパスや中日辞書データによる当該分野の機械翻訳精度の向上を図ることをその目的の一つとしている。

本品質評価は、主として以下の二段階にて実施する。

- (1) 機械的手法・書誌情報による人手品質評価対象分野の選定（一次評価）
- (2) 人手による品質評価の実施（二次評価）

また、上記品質評価に加え、昨年度事業における品質評価の結果、相対的に翻訳精度が低いとみなされ和文抄録作成対象とされた各テーマコードについて、当時の機械翻訳結果と、昨年度事業にて作成された中日辞書データの適用を含むバージョンアップが施された最新の機械翻訳結果とを人手で比較評価した。

- (3) 人手による新旧機械翻訳結果の比較評価の実施（新旧 MT 比較評価）

以下、それぞれの評価の実施概要について記す。

1.1.1 機械的手法・書誌情報による人手品質評価対象分野の選定（一次評価）

特許庁より貸与された「中日機械翻訳文データ¹」から平成 27 年発行の中国公開特許公報の要約の機械翻訳データを抽出し、これと中国公報データから抽出した原文データ、およびその人手翻訳文である「中国公開特許公報（A）和文抄録データ」とを対応づけて、原文、機械翻訳文、参照翻訳文からなる「評価対象文」を文献単位で作成した。

¹ 「中韓文献翻訳・検索システム」で提供されている機械翻訳結果と同一内容のもの。

この評価対象文を、機械翻訳文の著名な自動評価指標である BLEU²及び RIBES³を用いて自動評価し、その結果を、同じく特許庁より貸与された「中国公開特許公報の機械付与テーマコードリスト」及び「技術分野（中 AU）とテーマコードの対応表」に基づき 326 の技術分野（中 AU）単位で集計し、中 AU 単位で自動評価スコアの優劣順位を取得した。

この優劣順位をベースに、以下の各観点も加味して、人手品質評価（二次評価）の対象とする中 AU、全 120 分類を決定した。

1.1.2 人手による品質評価の実施（二次評価）

「1.1.1 機械的手法・書誌情報による人手品質評価対象分野の選定（一次評価）」にて選定した 120 の中 AU を対象に、それぞれ評価対象文を 50 文ずつ「中日機械翻訳文データ」から抽出のうえ、「特許文献機械翻訳の品質評価手順⁴」に従って、2 名の評価者による「(a) 内容の伝達レベル」及び「(b) 重要技術用語の翻訳精度」の評価を実施し、翻訳品質が相対的に低い中 AU を特定した。

各中 AU の評価対象文は、まず平成 27 年度発行の中国公開特許公報（A）から評価対象文献を 50 文献ずつ選定し、選定した文献の「要約」または「発明の詳細な説明」に相当する部分の中から 1 文献につき 1 文ずつ中国語の文を抽出した。その後、抽出した中国語の文に対応する機械翻訳文を「中日機械翻訳文データ」から取得するとともに、正解となる基準翻訳文の作成と、重要技術用語の選定を行った。

人手評価は、基準翻訳文を作成した者ではない 2 名の評価者により重複評価を行い、その平均値を評価結果とした。また「重要技術用語の翻訳精度」の評価対象となる技術用語は、技術分野ごとに抽出した 50 文のセットそれぞれから 50 語ずつ、計 6,000 語を特許庁担当者と相談の上選定した。

² 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.

³ 平尾努、磯崎秀樹、須藤克仁、Duh Kevin、塚田元、永田昌明：語順に基づく機械翻訳の自動評価法、自然言語処理、Vol.21, No.3, pp.421-444, 2014.

⁴ http://www.jpo.go.jp/shiryou/toushin/chousa/tokkyohonyaku_hyouka.htm

1.1.3 人手による新旧機械翻訳結果の比較評価の実施（新旧 MT 比較評価）

特許庁より貸与された「平成 28 年度品質評価対象テーマコードリスト」に記載された 174 テーマ、すなわち、昨年度事業にて機械翻訳精度が相対的に低いと見なされ、和文抄録作成対象とされた各テーマコードについて、上記 1.1.2 と同様の手法で計 500 文の評価対象文を選定した。さらに、この 500 文について、特許庁にて再翻訳された新たな機械翻訳文も取得し、新旧の機械翻訳文を揃えたうえで、2 名の評価者による「(a) 内容の伝達レベル」及び「(b) 重要技術用語の翻訳精度」の評価を実施し、新旧の機械翻訳文の品質の比較評価を行った。

1.2 一次評価の実施

一次評価、すなわち「機械的手法・書誌情報による人手品質評価対象分野の選定」では、平成 27 年の中国特許公開公報の「要約」（約 75 万件）の機械翻訳結果を対象とし、これを中 AU（326 分類）単位に区分したうえで、主要な自動評価指標である BLEU および RIBES を用いて自動評価を行い、中 AU レベルでの翻訳品質を比較評価した。以下、その詳細を示す。

1.2.1 一次評価対象データ

一次評価は、平成 27 年の中国特許公開公報の要約及び発明の名称の特許庁による機械翻訳結果を対象とした。なおリファレンス（正解文）には、過去の特許庁事業にて作成した人手翻訳文（和文抄録）を使用した。なお評価データは、各案件とも「要約」全文を 1 レコードとした。

一次評価の対象案件は、以下の各規準に則り、753,030 件とした。

- ① 平成 27 年発行の中国特許公開公報である：955,337 件
- ② 上記①のうち、和文抄録が存在する：814,875 件
- ③ 上記②のうち、貸与データ中に機械翻訳結果が存在する：810,437 件
- ④ 上記③のうち、「中国公開特許公報のテーマコード機械付与リスト」に存在し、テーマコードが付与されている：753,030 件

なお上記④は、評価対象の各案件の属する中 AU を定める際、機械付与されたテーマコードを用いる必要があることから定めた基準である。

1.2.2 中 AU の特定

一次評価では中 AU レベルで各分類の機械翻訳品質を比較評価する。このため、前述のとおり、評価対象の各案件について「中国公開特許公報のテーマコード機械付与リスト」を参照し、付与されているテーマコードから中 AU を特定のうえ、当該中 AU の評価データに加えることとした。

なお、一案件に対して複数のテーマコードが機械付与されている場合、その案件は、各テーマコードに対応する全ての中 AU の評価データに加えた。したがって、一案件が複数の中 AU の評価データに含まれる場合がある。

中 AU ごとの評価データの件数（文献単位）を別添 1-1「一次評価 中 AU 別対象データ

件数一覧」に示す。

1.2.3 BLEU 及び RIBES による自動評価結果

中 AU (全 326 分類) ごとの評価データについて、2 つの自動評価ツール BLEU と RIBES のそれぞれによる自動評価を行った。以下、その結果について分析する。

1.2.3.1 BLEU による自動評価結果

BLEU は広く一般的に使用されている自動評価ツールであり、基準翻訳文と機械翻訳文との間での N グラムの適合率に基づき評価を行う。用語の正確性に重みが置かれており、構文が似ている英語－仏語の間の翻訳等、翻訳時に大きな語順の入れ替えを必要としない言語間では、人手評価との相関が高くなる特徴を有する。特に用語の正確性は、本事業の品質評価においては極めて重要な要素であることから、自動評価指標の 1 つとして BLEU を採用した。

中 AU 別の BLEU 評価スコアの詳細は別添 1-2「一次評価結果（総合）及び二次評価対象中 AU 一覧」の「BLEU 評価結果」欄を参照されたい。また、BLEU 評価スコア順にプロットしたグラフを別添 1-3「BLEU 評価スコア順グラフ」として添付する。

本項では、BLEU による自動評価結果の評価順位上位 10 分類と下位 10 分類について下表 1-1 及び 1-2 に示す。

表 1-1 中 AU 別自動評価 (BLEU) 上位 10 分類一覧

順位	中 AU		データ数	平均スコア
1	JHV2-1	潤滑剤	2,522	0.376741
2	JJW1-1	高分子組成物	20,925	0.361819
3	JHV6-1	塗料・除去剤	9,668	0.359026
4	JHV4-1	化合物の物理・化学的用途	5,359	0.354922
5	GGG4-1	触媒	7,191	0.351512
6	JJJ3-1	付加系高分子（特殊）	1,872	0.347836
7	GGG3-1	金属化合物	4,260	0.345580
8	JHV7-1	接着方法・特殊接着剤・	3,149	0.343236
9	AGG4-1	M14 ナノ構造物	4,512	0.342819
10	JJJ5-1	縮合系高分子（熱硬化	4,912	0.342706

表 1-2 中 AU 別自動評価 (BLEU) 下位 10 分類一覧

順位	中 AU		データ数	平均スコア
317	DJW4-1	固着	3,211	0.188210
318	EBB3-1	紙送り	3,700	0.188027
319	EFF3-1	破碎・粉碎	2,607	0.187870
320	EBN2-2	容器	3,777	0.187363
321	EBN1-1	包装・紙製品製造	5,517	0.186458
322	CER4-1	家具	4,015	0.185944
323	EFS1-1	自動組立	2,013	0.185646
324	ECC5-1	携帯工具	5,986	0.185397
325	CBB3-2	水産	1,415	0.183374
326	EBN2-1	容器の蓋	1,986	0.176544

2つの表を参照すると、上位 10 位（翻訳品質が優れた分類）に関しては J で始まる中 AU (JH×4、JJ×3) が全 7 件を占め、化学物質／高分子系が目立つ結果となった。一方、下位 10 位（翻訳品質が劣る分類）では E で始まるものが 7 件 (EB×4、EF×2、EC) と、加工・製造技術に大きく偏る結果となった。

続いて、全 326 分類のデータ数及び平均スコアの平均値と、上位 10 分類、下位 10 分類限定での平均値とを下表 1-3 に示す。

表 1-3 中 AU 別自動評価 (BLEU) 平均値比較

	データ数範囲	データ数平均	平均スコアの平均
326 分類平均	21～64,316	4,431	0.249492
上位 10 分類平均	1,872～20,925	6,437	0.352620
下位 10 分類平均	1,425～5,986	3,423	0.185483

評価データの平均件数を比較すると、上位 10 分類が 6,437 件と、下位 10 分類の 3,423 件に比べて倍近い件数となっているが、これは第 2 位の JJW1-1 が 20,925 件とかなり突出したデータ数であるためであり、この分類を除いた 9 分類の平均は 4,827 件と、全体平均と大きな違いはない。そう考えると、上位下位どちらも、全分類の件数範囲や平均件数から大きな逸脱は見られず、穏当な数字であるといえる。したがって、今回の BLEU 評価においては、データ件数の多寡による影響はあまり考慮する必要はないと考えられる。

1.2.3.2 RIBES による自動評価結果

本事業では、BLEU と同一の評価データに対して、もう 1 つの代表的な自動評価指標である RIBES による評価も実施した。

RIBES は、BLEU の問題点である「翻訳時に大きな語順の入れ替えが必要となる言語対において、人手評価の相関が低くなる」という課題に対応すべく考案された自動評価ツールであり、基準翻訳文と機械翻訳文との間でのユニグラムの適合率に基づき評価を行う。特徴としては、訳語の違いを重視せず、語順を重視することで、人手評価との相関を高くしている。実際、NTCIR-7 での日英・英日機械翻訳タスクにおいては、人手評価とのシステム・レベルの順位相関（Spearman の ρ ）では、妥当性、流暢性ともに BLEU を明らかに上回っている。本品質評価の対象となる言語対は中国語－日本語であり、BLEU が苦手とする「翻訳時に大きな語順の入れ替えが必要となる言語対」に該当する可能性がある。このため、BLEU とは異なる手法による RIBES による評価を別途実施し、双方の結果を総合することで、より正確な評価結果を得ることとした。

中 AU 別の RIBES 評価スコアの詳細は別添 1-2「一次評価結果（総合）及び二次評価対象中 AU 一覧」の「BLEU 評価結果」欄を参照されたい。また、BLEU 評価スコア順にプロットしたグラフを別添 1-4「RIBES 評価スコア順グラフ」として添付する。

RIBES による自動評価結果の評価順位上位 10 分類と下位 10 分類について下表 1-4 及び 1-5 に示す。

表 1-4 中 AU 別自動評価（RIBES）上位 10 分類一覧

順位	中 AU		データ数	平均スコア
1	JHV2-1	潤滑剤(1)	2,522	0.830983
2	JHV6-1	塗料・除去剤(3)	9,668	0.828319
3	JJW1-1	高分子組成物(2)	20,925	0.823841
4	JJJ3-1	付加系高分子（特殊）(6)	1,872	0.818878
5	JJJ5-1	縮合系高分子（熱硬化(10)	4,912	0.817779
6	JHV7-1	接着方法・特殊接着剤・(8)	3,149	0.817084
7	JHV8-1	顔料、CB、木材スレイン(11)	2,129	0.814652
8	JJW2-1	高分子処理(13)	4,701	0.810165
9	GKK5-1	合金(21)	9,236	0.805272
10	JHV4-1	化合物の物理・化学的用途(4)	5,359	0.805208

（赤字カッコは BLEU 評価順位）

表 1-5 中 AU 別自動評価 (RIBES) 下位 10 分類一覧

順位	中 AU		データ数	平均スコア
317	EFF3-1	破碎・粉碎(319)	2,607	0.712749
318	CER4-1	家具(322)	4,015	0.712748
319	ECC5-1	携帯工具(324)	5,986	0.712072
320	EFS4-1	実装(284)	280	0.711914
321	CER2-1	錠(289)	4,504	0.711810
322	EBN1-1	包装・紙製品製造(321)	5,517	0.711542
323	EFS1-1	自動組立(323)	2,013	0.709509
324	EBB3-1	紙送り(318)	3,700	0.709411
325	EBN2-1	容器の蓋(326)	1,986	0.709148
326	CBB3-2	水産(325)	1,415	0.698147

(赤字カッコは BLEU 評価順位)

上位、下位 10 位の表を見ると、まず上位 10 位（翻訳品質が優れた分類）においては、J で始まる中 AU（JH×5、JJ×4）が 9 件を占め、化学物質／高分子系に大きく偏る結果となった。化学物質／高分子系は BLEU 評価でも 7 件を占めており、その全てが RIBES でも上位 10 位に入っている。なお、化学物質／高分子系の残る 2 分類（JHV8-1、JJW2-1）及び第 9 位の GKK5-1（合金）も、全て BLEU 評価の上位 21 位以内に入っており、少なくとも品質の優れた分野に関しては、BLEU と RIBES とで極めて近似した結果が得られたといえる。

一方、下位 10 位に関しては、E で始まる中 AU（加工・製造技術関連）が 7 件（EB×3、EF×3、EC）を占めており、こちらも BLEU と同様の傾向を示した。これら 7 件中の 6 件を含む 8 件が、BLEU 評価でも下位 10 位内に入っており、残る 2 分類（EFS4-1、CER2-1）も、それぞれ BLEU 評価は 284 位、289 位と低い順位となっており、全体的な傾向としては、上位 10 位と同様、BLEU と酷似しているといえる。

最後に、RIBES 評価結果についても、全 326 分類のデータ数及び平均スコアの平均値と、上位 10 分類、下位 10 分類限定での平均値とを比較した。結果を下表 1-6 に示す。

表 1-6 中 AU 別自動評価 (RIBES) 平均値比較

	データ数範囲	データ数平均	平均スコアの平均
326 分類平均	21～64,316	4,431	0.749275
上位 10 分類平均	1,872～20,925	6,447	0.817218
下位 10 分類平均	280～5,986	3,202	0.709905

RIBES も、上位 10 分類と下位 10 分類との平均データ件数に顕著な開きがあり、その差は BLEU よりもやや開いている。ただし、これは、上位 10 分類において BLEU と同じく JJW1-1 (データ件数 20,925 件) という件数の大きい分類が含まれており、加えて下位 10 分類において、BLEU では下位 10 位に含まれていなかった極端に件数の少ない EFS4-1 (280 件) が含まれたことで平均データ件数がさらに小さくなり、上位下位の差が開いたためである。これらの「特異値」といえる 1 分類を除いた 9 分類の平均件数はそれぞれ 4,839 件、3,527 件と穏当な数値となり、RIBES 評価においても、BLEU 評価と同様、データ件数の多寡と評価結果との明確な関連性は見られないと結論できる。

1.2.3.3 BLEU と RIBES の評価結果の比較

1.2.3.2 で述べたとおり、評価順位上位、下位いずれにおいても、BLEU と RIBES の評価値は極めて近い結果であることが見て取れた。

これをさらに検証するため、BLEU と RIBES の評価値順位の差がどの程度の範囲に分布するかを、中 AU 全 326 分類について調査した。評価順位の差とは、例えば BLEU の評価順位が 10 位で、RIBES の評価順位が 15 位の場合、差は 5 となる。なお BLEU と RIBES のどちらが上位であっても、差は絶対値（常に正の値）とした。

下表 1-7 に、中 AU 全 326 分類における BLEU と RIBES の評価順位差の分布を示す。

表 1-7 中 AU 別 BLEU/RIBES 評価順位差分布

評価順位差	分類数	構成比
0 ～ 25	272	83.4%
26 ～ 50	48	14.7%
51 ～ 75	5	1.5%
76 ～ 100	1	0.3%
101 ～ 326	0	—

上表のとおり、2 つの自動評価結果の順位差が 25 位以内のものが 272 分類 (83.4%) と大勢を占め、50 位以内まで含めると 98.1% と、ほぼ全ての分類がカバーされる。一方、51 位以上の差となったものは 6 件のみとごく僅かであった。このことから、BLEU と RIBES の評価は、少なくとも中 AU 別の相対評価という観点からは極めて高い整合性を有しており、どちらを使用してもほぼ同様の相対順位が得られるといえる。

なお、2 つの評価順位の差が 51 位以上と大幅に異なった分類は下表 1-8 に示す 6 分類である。

表 1-8 BLEU/RIBES で評価順位差が 51 位以上であった中 AU 一覧

中 AU		BLEU 順位	RIBES 順位	順位差
CBB4-1	栽培、捕獲・駆除	195	113	82
ZEU2-1	メモリ回路・信頼性	122	187	65
EA21-1	鑄造	181	120	61
ZKX3-1	増幅器	72	130	58
HDD1-1	固体廃棄物	136	85	51
ZCC1-1	ビデオ記録	117	168	51

6 件中、BLEU の評価のほうが高かったもの（＝順位の小さかったもの）が 3 件（ZEU2-1、ZKX3-1、ZCC1-1）、RIBES のほうが高かったものも 3 件（CBB4-1、EA21-1、HDD1-1）であった。これらの分類について BLEU と RIBES のどちらの評価がより妥当かは検証が難しいが、これらの 6 分類については、二次評価の対象に加え、二次評価（人手評価）の結果により、それぞれの分類において、BLEU と RIBES のどちらが妥当かを別途検証してみる。

1.2.3.4 一次評価スコアの算出

一次評価では、中 AU 全 326 分類に対し、BLEU と RIBES それぞれの評価スコアを得た。一次評価の最終的な結果を得るためには、これら各種評価スコアを総合する必要がある。

このため、1.2.3.1～1.2.3.3 にて BLEU と RIBES それぞれの評価スコアの傾向を分析した。その結果、BLEU と RIBES の評価結果は、ごく僅かな例外はあるものの、全体的には非常に高い相関性があることが確認された。

このことは、総合評価スコアの算出にあたって、BLEU と RIBES それぞれの評価スコアに一定の重みづけを施したとしても、総合順位には大きな変動は生じないことを意味する。また、そもそも双方の評価スコアに顕著な差異が見られないため、それぞれの重みづけを定める根拠に乏しいのも事実である。

そこで、一次評価については、各評価スコアに一切の重みづけを行わず、BLEU と RIBES のそれぞれについて、評価データ全件の平均スコアに基づく評価順位を取得し、さらに、2 つの評価指標それぞれの評価順位を 1 : 1 で加算した総合順位に基づいて、一次評価の最終評価結果とすることとした。

別添 1-2「一次評価結果（総合）及び二次評価対象中 AU 一覧」に、上記集計方法に基づく、一次評価の最終評価結果を示す。なお、この一次評価は、二次評価の対象とする品質下位 120 分類を決定するためのものであるため、本資料は、総合評価の低い順（相対的な翻訳品質の悪い順）に並べている。

なお、1.2.3.3 に示した、2つの評価指標の順位差が 51 位以上であった 6 分類については、先述のとおり、二次評価の結果を踏まえ精査する意義が大きいため、無条件で二次評価の対象とすることとした。

それぞれの分類の、一次評価最終評価結果における順位は下表 1-9 のとおりである。結果、6 分類について、総合評価下位の分類と差し替えて、二次評価対象に追加されることとなった。

表 1-9 特異評価値の中 AU の総合評価順位（評価の低い順）

	中 AU		総合順位
BLEU と RIBES での順位差 51 以上の 6 分類	CBB4-1	栽培、捕獲・駆除	170
	ZEU2-1	メモリ回路・信頼性	169
	EA21-1	鋳造	175
	ZKX3-1	増幅器	224
	HDD1-1	固体廃棄物	215
	ZCC1-1	ビデオ記録	186

1.2.4 補足的な評価観点の反映

二次評価の対象とすべき中 AU の選定にあたり、一次評価の総合評価結果とあわせ、以下の各観点からの検討も実施した。

- ①日本特許庁における審査件数（相対的に多いもの）
- ②日本特許庁における審査請求率（相対的に高いもの）
- ③日本語ファミリー有り文献率（相対的に低いもの）

以下、それぞれの観点についての分析結果を記す。

1.2.4.1 日本特許庁における審査件数

日本特許庁において中国文献の審査件数が多い中 AU は、日本国において特に重要な技術分野とみなせるため、この分類に属する中国特許文献の読解へのニーズも高いと考えるのが妥当である。このため、こうした分野の機械翻訳品質については特に入念に評価する必要がある、一次評価の結果によらず、二次評価の対象に加えることも検討すべきである。

この考えのもと、まずは本事業の貸与物である「テーマコード情報表」に基づき、2011 年～2013 年におけるテーマコード別の日本特許庁への審査請求件数を、中 AU 単位に集計した。その概要は以下のとおりである。

- ・ 全中 AU（326 分類） 平均：2,131 件
- ・ 審査請求件数最多：CBN5-1「パチンコ」18,833 件
- ・ 審査請求件数最少：EFS2-2「特殊目的用コンベア」52 件

全中 AU の審査件数順集計結果は別添 1-5「中 AU 別 日本特許庁における審査件数一覧」を参照されたい。

続いて、審査件数の順位と、一次評価結果の順位との相関を調査した。下表 1-10 に、審査件数上位 10 位の中 AU と、それぞれの一次評価総合順位（下位順）を示す。

表 1-10 審査件数上位 10 位の中 AU と一次評価総合順位（下位順）

順位	中 AU		審査件数	一次下位順
1	CBN5-1	パチンコ	18,833	13
2	YJJ2-1	移動無線通信システム	12,728	204
3	ZEE2-1	MM I	8,876	181
4	YLL2-1	電子商取引	8,595	158
5	ZGT1-1	電力系統	7,618	191
6	JJW1-1	高分子組成物	7,584	325
7	GKX1-1	電池・電極	6,800	301
8	DHV1-1	電動機・発電機	6,457	103
9	HBB1-1	遺伝子・蛋白質工学	6,095	278
10	HBN3-1	発酵・細胞工学	5,919	271

審査件数上位 10 位中、一次評価総合順位が 120 位以内（＝品質下位につき二次評価対象）であったのは 2 分類のみであり、残り 8 分類中 5 分類は 200 位以下と、後半の順位（＝品質上位）に偏った。中国案件の審査件数が多いということは、それだけ機械翻訳の精度向上に利用できる言語資源が豊富であるということであり、この結果は妥当といえる。

そう考えると、「審査件数が特に多い中 AU は日本における重要度が特に高い分類であるといえるため、自動評価値が高くても、二次評価の対象に加える」というのがこの観点の主旨ではあるものの、最終的には言語資源の充実を図るべき技術分野を決定するための二次評価に、これら 8 分類全てを追加する必要性は低い。そこで、最低限の対応として、審査件数が最上位に近い 2 位～5 位の 4 分類のみを二次評価の対象とすることとした。

1.2.4.2 日本特許庁における審査請求率

日本国において審査請求率が高い中 AU は、審査件数が多いものと同様、他の中 AU に比べて重要度は相対的に高いと考えられる。このため、審査請求率が特に高い中 AU について、一次評価の結果に拠らず二次評価対象に加えることを検討した。

まずは「テーマコード情報表」に基づき、2011 年～2013 年におけるテーマコード別の日本特許庁への審査請求率を中 AU 単位に集計した。その概要は以下のとおりである。

- ・全中 AU (326 分類) 平均 : 69.39%
- ・審査請求率最高 : CBN5-1 「パチンコ」 93.55%
- ・審査請求率最低 : ZDD2-1 「光記録」 44.41%

全中 AU の審査請求率順集計結果は別添 1-6 「中 AU 別 日本特許庁における審査請求率一覧」を参照されたい。

審査請求件数に続き、CBN5-1 が最上位となったが、それ以降は、請求件数と請求率とに目立った関連性はない。また、請求率最低の ZDD2-1 でも 44.41%と極端に低率というわけではなく、件数に比べると各中 AU 間の差は小さい。

審査請求率の順位と、一次評価結果の順位との相関についても調査した。下表 1-11 に、審査請求率上位 10 位の中 AU と、それぞれの一次評価総合順位（下位順）を示す。

表 1-11 審査請求率上位 10 位の中 AU と一次評価総合順位（下位順）

順位	中 AU		請求率	(審査数)	一次
1	CBN5-1	パチンコ	93.55%	18,833(1)	13(重複)
2	CBN5-2	スロット	92.01%	5,099(22)	16
3	CDD6-1	トンネル	87.77%	782(268)	49
4	ECC2-1	工作機械の制御	87.16%	584(286)	183
5	EFS4-1	実装	83.97%	1,100(234)	25
6	YJJ2-1	移動無線通信システム	83.05%	12,728(2)	204(重複)
7	CBN2-1	ゲーム・玩具	82.61%	2,561(92)	96
8	ZKK1-1	符号変換	82.15%	3,235(56)	238
9	FEE3-1	処置具	82.05%	3,794(38)	76
10	CDD3-1	基礎工	81.24%	2,083(126)	48

こちらは審査件数のときと大きく異なり、上位 10 位中、一次評価総合順位（下位順）が

120 位以内の中 AU が 7 分類も存在する結果となった。審査請求率の場合、率が高くても必ずしも件数が多いとは限らないが、表に「(審査数)」として示したように、大半の分類は審査件数もそれなりに上位であり、1.2.4.1 で述べた考察とはやや矛盾する結果となった。

審査請求率が高いほど機械翻訳品質が低くなるという傾向は合理的に説明がつくものではないが、結果的には、審査請求率の最も高い中 AU10 分類の大半は既に二次評価の対象となっており、「審査請求率が高い」＝「日本において重要性が高い」中 AU について二次評価で精査するという目的には適っている。ここでは、上位 10 位中、二次評価対象外となっている ECC2-1 と ZKK1-1 の 2 分類（※第 6 位の YJJ2-1 は 1.2.4.1 にて既に追加されているため不要）について、二次評価対象に追加することとした。

1.2.4.3 日本語ファミリー有り文献率

現状、中日機械翻訳の精度向上のための言語資源は、主に中国文献と日本語ファミリー文献の対訳データをソースとしている。このため、日本語ファミリーの存在する率が低い分野は、他の分野に比べて言語資源に乏しく、したがって機械翻訳品質が劣るという可能性がある。このため、これに該当する中 AU は、無条件で二次評価対象に加え、人手評価による精密な評価が必要であると考ええる。

この観点から、「テーマコード情報表」に基づき、中国特許文献数に対する日本語ファミリー文献数の比率を中 AU 単位に集計した。その概要は以下のとおりである。

- ・ 全中 AU (326 分類) 平均 : 11.04%
- ・ 審査請求率最低 : CDD6-1 「トンネル」 0.56%
- ・ 審査請求率最高 : BCC2-1 「電子写真・制御」 70.61%

全中 AU の日本語ファミリー率順の集計結果は別添 1-7 「中 AU 別 日本語文献ファミリー率一覧」を参照されたい。ファミリー率は審査請求率と異なり、中 AU ごとにかかなり大きな差がある (0.56%~70.61%) ことがわかる。

日本語ファミリー率の順位 (低率順) と、一次評価結果の順位との相関についても調査した。下表 1-12 に、日本語ファミリー率上位 10 位の中 AU と、それぞれの一次評価総合順位 (下位順) を示す。

表 1-12 日本語ファミリー率低率順 10 位の中 AU と一次評価総合順位（下位順）

順位	中 AU		ファミリー率	(ファミリー数)	一次
1	CDD6-1	トンネル	0.56%	27	49(重複)
2	CBB4-3	バイオ・その他	0.87%	167	284
3	HCC3-2	動植物物質含有医薬	0.90%	422	220
4	CBB4-2	木材加工	0.95%	39	252
5	JHV5-1	化合物のその他の用途	0.99%	147	298
6	CDD2-1	水工	1.07%	81	61
7	CDD3-1	基礎工	1.18%	102	48 (重複)
8	CBB4-1	栽培、捕獲・駆除	1.23%	330	170
9	CEE1-1	建築構造・部材	1.28%	139	71
10	HBN1-1	食料品	1.34%	1,209	247

本来、ファミリー率の低い分類は言語資源に乏しく、機械翻訳品質も相対的に低いことが想定されるが、一次評価結果を見る限り、下位 120 位に入る分類は 10 分類中 4 分類にとどまり、想定したような傾向は特に見られない。とはいえ、本項の意図は「ファミリー率の低い分類は機械翻訳品質が劣る可能性が高いため、一次評価の結果に関わらず二次評価に追加する」というものであり、昨年度も、ファミリー率 1%以下の分類は無条件で二次評価の対象としていた。本年度もこれに倣い、ファミリー率 1%以下かつ現時点で二次評価対象外である 4 分類（上表 2～4 位）を、二次評価対象に加えることとした。

1.2.4.4 昨年度品質評価結果との比較

昨年度事業（平成 28 年度「中国特許文献の解析及びデータ作成事業」）にて実施した翻訳品質評価では、一次評価として、全中 AU（326 分類）に対し評価文を 10 文ずつ用意し、人手評価を実施している。

そこで、当時の人手評価結果に基づく中 AU の順位（下位順）と、今回の一次評価の各種スコア（総合評価／BLEU 評価／RIBES 評価）とを比較した。下表 1-13 に、双方の評価順位差の分布を示す。なお、中 AU 別の昨年度評価結果と本年度評価スコアの対比表は別添 1-8「中 AU 別 昨年度一次評価結果と本年度一次評価結果の対比表」を参照のこと。

表 1-13 中 AU 別 昨年度評価結果と今回評価との順位差分布

評価順位差	総合		BLEU		RIBES	
	分類数	構成比	分類数	構成比	分類数	構成比
0 ～ 25	80	24.5%	81	24.8%	82	25.2%
26 ～ 50	58	17.8%	56	17.2%	62	19.0%
51 ～ 75	49	15.0%	55	16.9%	40	12.3%
76 ～ 100	37	11.3%	32	9.8%	35	10.7%
101 ～ 125	23	7.1%	25	7.7%	28	8.6%
126 ～ 150	27	8.3%	23	7.1%	27	8.3%
151 ～ 175	13	4.0%	19	5.8%	10	3.1%
176 ～ 200	14	4.3%	16	4.9%	13	4.0%
201 ～ 225	17	5.2%	8	2.5%	19	5.8%
226 ～ 250	4	1.2%	7	2.1%	5	1.5%
251 ～ 275	2	0.6%	2	0.6%	3	0.9%
276 ～ 300	2	0.6%	2	0.6%	2	0.6%
301 ～ 326	0	—	0	—	0	—

上表のとおり、昨年度の評価順位と今回の評価順位とは、順位差が 100 位以内に収まるものが約 7 割と一定の整合性は見られるが、101 位以上の大きな順位差が生じた中 AU も 3 割程度存在するなど、両者で矛盾するケースも存在した。ただし、昨年度と本年度では評価対象としたデータが完全に異なり、かつ昨年度は 1 分類あたり 10 文という限られたデータ数での評価であったことを考えると、かなり高い整合性であるともいえる。

なお、昨年度評価結果との比較は、総合評価のほか、BLEU、RIBES それぞれの単独評価

結果についても行ったが、もともと BLEU と RIBES の評価結果が高い整合性を有していたため、比較結果もほぼ同じような傾向となっている。

いずれにせよ、二次評価対象中 AU の選定は基本的には今回の一次評価結果に基づき行うことになるが、昨年度評価結果を本年度の二次評価に反映するとすれば、例えば昨年度の一次評価の品質下位 5～10 分類程度を念のため本年度も人手評価することが考えられる。

この考えから、昨年度評価結果下位 11 分類（9 位が同スコアで 3 分類につき）を対象に、今回の一次評価順位を調査した。結果を下表 1-14 に示す。

表 1-14 昨年度評価結果下位 11 分類の本年度一次評価順位（下位順）

昨年度順位	中 AU		一次
1	DGT1-1	タービン	148
2	ECP2-1	非金属の加工	14
3	DHH3-1	交通制御	203
4	CBN5-1	パチンコ	13
5	ECP7-1	装飾技術	90
6	DDD6-1	武器	106
7	DJW4-1	固着	10
8	FKK1-1	コネクタ	43
9	CBN1-1	遊具	58
9	EFF3-1	破碎・粉碎	9
9	ZBN2-1	ケーブル処理	42

調査の結果、上表に示したとおり、昨年度順位のの下位 11 分類中 9 分類については、本年度一次評価でも下位 120 分類中に含まれており、自動的に二次評価の対象となることが判明した。この結果は、昨年度の一次評価が高い精度を有することを示しているといえる。したがって、残る 2 分類（DGT1-1、DHH3-1）についても、念のため二次評価に加え、再度の精査を行うのが妥当であると判断した。

1.2.5 二次評価対象中 AU の決定

二次評価の対象とする中 AU は 120 分類である。基本的には一次評価の総合評価順位の
下位順とするが、下表 1-15 にまとめた各分類（全 19 分類）については、各々の理由により
優先的に対象に追加した。

表 1-15 二次評価にて優先的に対象とする中 AU 一覧

No.		中 AU		
1	①BLEU と RIBES の評価順位差が 51 位以上の 6 分類 (1.2.3.3 参照)	CBB4-1	栽培、捕獲・駆除	170
2		ZEU2-1	メモリ回路・信頼性	169
3		EA21-1	鋳造	175
4		ZKX3-1	増幅器	224
5		HDD1-1	固体廃棄物	215
6		ZCC1-1	ビデオ記録	186
	②日本における 審査件数 上位 5 分類 (1.2.4.1 参照)	CBN5-1	パチンコ	13
7		YJJ2-1	移動無線通信システム	204
8		ZEE2-1	MM I	181
9		YLL2-1	電子商取引	158
10		ZGT1-1	電力系統	191
	③日本における 審査請求率 上位 10 分類 (1.2.4.2 参照)	CBN5-1	パチンコ	13(重複)
		CBN5-2	スロット	16
		CDD6-1	トンネル	49
11		ECC2-1	工作機械の制御	183
		EFS4-1	実装	25
		YJJ2-1	移動無線通信システム	204(重複)
		CBN2-1	ゲーム・玩具	96
12		ZKK1-1	符号変換	238
		FEE3-1	処置具	76
		CDD3-1	基礎工	48
	④日本語ファミリ ー率が 1%以下 全 5 分類 (1.2.4.3 参照)	CDD6-1	トンネル	49(重複)
13		CBB4-3	バイオ・その他	284
14		HCC3-2	動植物物質含有医薬	220
15		CBB4-2	木材加工	252
16		JHV5-1	化合物のその他の用途	298
17	⑤昨年度評価 品質下位	DGT1-1	タービン	148
		ECP2-1	非金属の加工	14

18	全 11 分類 (1.2.4.4 参照)	DHH3-1	交通制御	203
		CBN5-1	パチンコ	13(重複)
		ECP7-1	装飾技術	90
19		DDD6-1	武器	106
		DJW4-1	固着	10
		FKK1-1	コネクタ	43
		CBN1-1	遊具	58
		EFF3-1	破碎・粉砕	9
		ZBN2-1	ケーブル処理	42

上表のとおり、優先的に二次評価対象とする中 AU が 19 分類存在するため、一次評価結果からは品質下位順に 101 位までを対象とした。

二次評価の対象とする中 AU、全 120 分類の内訳については、[別添 1-2「一次評価結果（総合）及び二次評価対象中 AU 一覧」](#)の「二次評価対象」欄を参照されたい。

1.3 二次評価の実施

二次評価では、一次評価にて選定した人手品質評価対象分野、すなわち 120 種の中 AU に対して、人手による機械翻訳品質の評価を実施した。具体的には、各中 AU について評価対象文を 50 文ずつ用意し、「特許文献機械翻訳の品質評価手順」に従って、2 名の評価者による「(a) 内容の伝達レベル」及び「(b) 重要技術用語の翻訳精度」の評価を実施した。以下、その詳細を示す。

1.3.1 評価対象文献の選定

二次評価の対象文献は、平成 27 年発行の中国公開特許公報について、特許庁から貸与された「中国公開特許公報の機械付与テーマコードリスト（平成 15～27 年発行の中国公開特許公報に対しテーマコードを機械的に付与したテーマコードリスト）」及び「技術分野（中 AU）とテーマコードの対応表」に基づき、各文献の属する中 AU を特定したうえで、二次評価対象となる 120 の中 AU のそれぞれに属する文献を 50 件ずつ選定した。

なお、各中 AU それぞれについて評価対象文献 50 件を選定するにあたり、この 50 文献におけるテーマコード（中 AU の下層分類）を、実際の中国文献における分布を反映した構成とした。具体的には、以下の基準で 50 文献を選定した。

- ① 中 AU ごとの中国公開特許文献のテーマコード別分布を調査
- ② 各中 AU におけるテーマコードの分布率に基づき 10 文献の配分を決定

つまり、ある中 AU に属する中国公開公報全件におけるテーマコード構成が、テーマコード A が 50%、B が 30%、C と D が 10%であれば、選定する 50 文献の構成は、テーマ A に属する文献を 25 文献、テーマ B を 15 文献、テーマ C と D を 5 文献ずつとした。このような基準でサンプリングを行うことで、それぞれの中 AU に属する主なテーマコードを全てカバーし、かつ各中 AU におけるテーマコードの分布構成も反映した形で評価対象文献を選定することができる。

次表に、上記構成の実例として中 AU 「CBB3-2」における 50 文献の構成を示す。

表 1-16 中 AU「CBB3-2」の二次評価対象 50 文献の構成

中 AU	テーマコード	中国文献数（構成比）	50 文献の構成
CBB3-2	2B307	469（23.1%）	11 文献
	2B109	441（21.7%）	11 文献
	2B106	437(21.5%)	11 文献
	2B105	357（17.6%）	9 文献
	2B108	198（9.8%）	5 文献
	2B019	128（6.3%）	3 文献

1.3.2 評価対象文の選定

上記 1.3.1 にて選定した二次評価対象文献 6,000 文献（120 中 AU×50 文献）について、1 文献あたり 1 文ずつ、評価対象文を選定した。以下、評価対象文の選定基準について記す。

1.3.2.1 選定基準

本品質評価の目的は、「中韓文献翻訳・検索システム」を通じて提供している中国特許の機械翻訳データについて、相対的に翻訳精度の低い技術分野を特定することである。

通常、機械翻訳の品質評価は、同一の入力文を用いて、複数の異なるエンジンや、バージョンアップ前後のエンジンにおける出力結果を比較するという形式が一般的であるが、本品質評価はこれと異なり、「単一のエンジンに対し、技術分野ごとに異なる入力文を用いてその機械翻訳結果を比較し、技術分野間の優劣をつける」必要がある。したがって、評価を正確に行うポイントは、各技術分野の評価対象文を、いかに「本来、同等の翻訳難易度と考えられる文」に揃え、各分野の条件を公平にできるかにある。

とはいえ、自然文の「本来の翻訳難易度」を測定することは極めて困難であり、明確な基準は存在しない。このため、本評価においては、各中 AU にて評価対象となる 50 文について、以下の厳密な選定基準を設け、原則、全ての文をこの条件で一律に揃えることで、技術分野間の難易度のばらつきを極力抑えることとした。

- A. 対象文の文長を一定範囲（30～50 文字）に揃える
- B. 対象文の抽出箇所を「発明の詳細な説明」に揃える
- C. 難易度を左右する特殊な表現を含む文は除外する
- D. 所定文字数（3～6 文字）の重要技術用語候補を含む文のみを対象とする

以下、各条件について説明する。

1.3.2.2 選定基準 A：対象文の文長を一定範囲（30～50 文字）に揃える

機械翻訳の精度は、基本的に入力文が長くなるにつれ悪化する。したがって、技術分野ごとの評価対象文の文長がばらついてしまうと、文長の差の影響が強くなり、各分野の翻訳品質の公平な比較にならない。このため、二次評価においては、中 AU ごとに抽出する 50 文の文長を一定の範囲内に限定し、各中 AU 間の比較条件を極力揃えることとした。

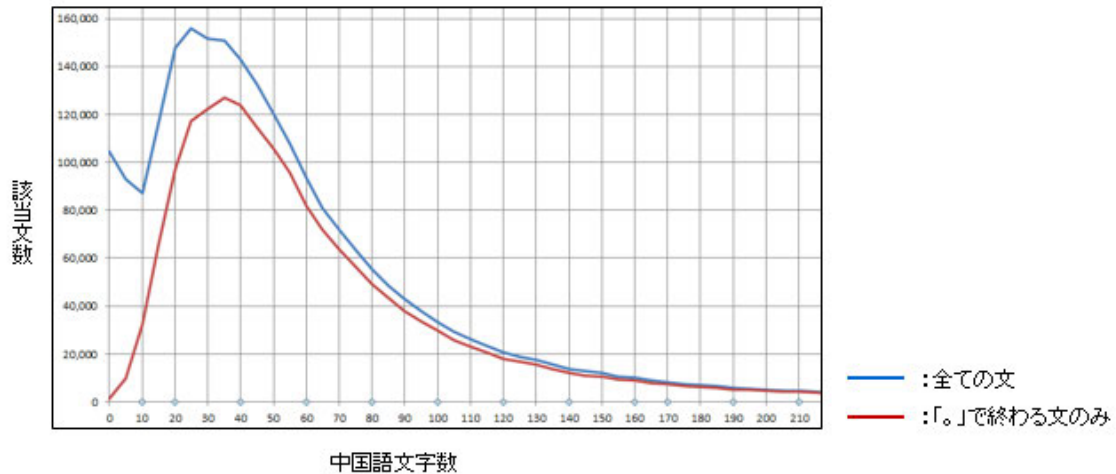
なお、中 AU 間の条件を一律にするという意味では、複数の文長範囲を設定し、例えば 15 文、20 文、15 文など、50 文をそれぞれに割り振る手法も考えられる。しかしながら、本評価で最も重要なポイントは「各分野の『本来の翻訳難易度』を極力同等に揃えること」であり、かつ、この条件には明確な基準が存在せず、文長などコントロール可能な諸条件を可能な限り均一としたとしても、依然として偶然の要素が少なからず残留することは避けられないという事情がある。こうした要素を極力軽減するためには、同一条件のサンプル数をなるべく増やすことが非常に重要となってくる。このことから、本評価においては、複数の異なる文長範囲の文を少しずつ比較するよりも、50 文すべてを同一の文長範囲とし、サンプル数を多くするほうが、より平均的なデータ集合となり、公平な評価となる確率を高めることができる判断に至った。

加えて、文長の違いによる翻訳品質の差の把握は、異なるエンジン間の優劣を測る場合には非常に重要であるが、本調査は同一のエンジンでの比較であり、かつ「異なる文を（本来）同程度の難易度に揃える」という難しい条件を達成する必要があることから、サンプル数を増やしてこの条件を達成することを優先した。

50 文に適用する文長範囲については、極端に短いものや長いものは、翻訳の難易度も極端に低く／高くなり、結果として分野間の評価に差が出にくくなるため避けるべきであるが、それ以外は理論上、特段の適不適はない。ただし、中国文献において最も出現頻度の多い文長範囲を採用すれば、豊富な候補の中から本調査に最適の文を選定することができるため、分野間の条件のばらつきを最小限に抑えるという観点から最良と考えた。

この考えのもと、まずは実際の中国特許文献における文長の分布（5 文字単位）を、中国公開特許明細書 1 ヶ月分（2015 年 12 月公開分）全件、約 180 万文について調査し、中国文献の文長構成に関する分野横断的な最新の傾向を把握した。下記グラフ（図 1-17）にその結果を示す。

図 1-17：中国公開特許公報（2015.12）における文長分布



このグラフでは、1 ヶ月分の中国公開公報の文長別の分布を、全ての文（青線）、および句点（。）で終わる文のみ（赤線）とで示しており、句点で終わる文（＝自然文とみなせる文）の文字数のピークは 35～40 文字の範囲であることがわかる。この結果から、中国特許文献全般においてはこのピーク周辺のものが最も文数が多く、したがってサンプルが豊富であるといえる。この文字範囲周辺は機械翻訳の難易度も適度な印象もあることから、本調査では、このピーク文字範囲 35～40 文字をコアとした 30～50 文字を評価対象文の文長の条件として、全ての評価対象文を極力この範囲のものに揃えることとした。

1.3.2.3 選定基準 B：対象文の抽出箇所を「発明の詳細な説明」に揃える

本品質評価では、対象文の抽出箇所が「要約」または「発明の詳細な説明」に相当する部分であることが要件となっているが、本評価では、各中 AU の評価対象文 50 文の全てを「発明の詳細な説明」から抽出した。

中国特許文献の「要約」では長文が多用される傾向にあり、本来であれば文を区切るべき話題の転換時にも、そのまま文が継続されがちである。中国文献の「要約」の平均文字数は 260 文字前後であるが、その多くは 1 文～3 文程度で構成されており、「文長が極端に長く、（翻訳難易度が極端に高くなるため）技術分野間の差が出にくい」文が大半であると考えられる。こうした限られた数文から本評価の基準である「30～50 文字」の文を搜索することは効率が悪く、条件が揃わない分野が多発するおそれがある。

一方、「発明の詳細な説明」は、候補となる文の数が圧倒的に多く、豊富なサンプル群から各分野の代表的な文を選定することができ、かつ文長をはじめとする選定文の諸条件を

各分野で公平に揃えられる可能性がより高い。また、「発明の詳細な説明」は権利請求の解釈の対象範囲であるため、「要約」よりも正確な文で記載されている可能性が高く、文法的・内容的に辻褃の合わない文を誤って対象文としてしまうリスクも低くなる。

本評価では、これらの理由により、評価対象文は全て「発明の詳細な説明」から抽出すべきと結論した。

1.3.2.4 選定基準 C：難易度を左右する特殊な表現を含む文は除外する

自然文の中には、機械翻訳が苦手とする特殊な表現を含むものがある。例えば、文の途中にカッコ書きで長い補記が挿入されているような文は、それだけで他の文よりも機械翻訳の難易度は高くなる。こうした文が、無作為に一部の技術分野のみで評価対象文として選定されると、その分野は他の分野よりも不当に不利となる。

また、当然ながら原文に誤記や、その出願人しか使わないような特殊な用語・表現が含まれるような文は、これが原因で翻訳精度が低下することは明白であり、対象文に含めるべきではない。

こうした観点から、本評価では、評価対象文からこのような特殊な文を排除することとした。以下、その例を列挙する。

- ・誤字やタイプミスを含むもの
- ・長いカッコ書き補記を含むもの
- ・英数字記号等を多く含むもの
- ・上付・下付文字などの特殊文字を含むもの
- ・出願人特有の造語や独特の表現を含むもの
- ・固有名詞（会社名や商品名など）を含むもの

1.3.2.5 選定基準 D：所定文字数(3～6)の重要技術用語候補を含む文のみを対象とする

評価対象文と同様、その中からピックアップする重要技術用語についても、技術分野間で不公平な評価とならぬように配慮する必要がある。

技術用語を機械翻訳する際、概して単語よりも複合語（複数の単語からなる語）のほうが難易度が高くなることは言うまでもない。したがって、技術分野によって選定される用語の長さが不揃いでは、分野間の条件に根本的な有利不利が生じ、正確な分野間比較とならない。

本評価では、重要技術用語の翻訳精度評価においてこうした事態となることを回避し、分

野間の翻訳精度の優劣が公平な条件で評価されるよう、選定する重要技術用語の文字数範囲を全中 AU において一律に揃えることとした。具体的には、各技術分野にて選定する全ての重要技術用語候補を、原則として、あらゆる分野で候補が豊富に存在し、かつ技術分野ごとの特徴が出やすいと考えられる、中国語 3～6 文字の用語（2～3 語からなる複合語が該当することが多いと思われる）に限定し、各分野の有利不利が極力解消されるようにした。

重要技術用語は、評価対象文のそれぞれから 1 語ずつ抽出することになる。このため、評価対象文の選定にあたっては、あらかじめ上記条件に該当する重要技術用語の候補を含んでおり、かつ技術分野で同一の語が重複しないことを文選定の条件とした。

1.3.2.6 二次評価対象文の詳細

上記の各基準に従い選定した二次評価対象文 6,000 文の詳細は下記電子ファイルを参照されたい。

『01 二次評価（中 AU）文単位評価結果.xlsx』

1.3.3 重要技術用語の選定

二次評価における「重要技術用語の翻訳精度」の評価のための重要技術用語の選定は、上記 1.3.2 にて選定した二次評価対象文 6,000 文それぞれから 1 語ずつ、計 6,000 語を選定した。二次評価対象文の選定にあたっては、前記 1.3.2.5 にて述べたとおり、あらかじめ所定の文字数範囲の技術用語を含む文を選んでおり、各文からこの技術用語を採用した形となる。

重要技術用語についても、各中 AU で難易度を揃える必要があることから、前述のとおり、本評価では全ての中 AU で選定する全ての重要技術用語を「中国語 3～6 文字」のものに揃えた。この条件により、本評価で選定される重要技術用語は、ほぼ全てが 2～3 語で構成される複合語に統一される。

なお、重要技術用語は、同一分野で同一の用語を重複して選定することは不可としたが、分野間では重複を認めた。これは、一つの用語は必ずしも一つの分野のみで使われるものではなく、各分野で無作為に抽出された用語が分野間で重複した場合は、その語はそれぞれの分野で重要な技術用語であるため、無用な調整はせず、それぞれの分野にて評価対象にすべきである、との考えに基づく。

二次評価における重要技術用語 6,000 語の詳細は下記電子ファイルを参照されたい。

『01 二次評価（中 AU）文単位評価結果.xlsx』

1.3.4 評価基準

二次評価では、各評価対象文に対して「内容の伝達レベル」と「重要技術用語の翻訳精度」の二つの観点から人手評価を行った。人手評価にあたっては、評価精度を高めるため、両観点とも、評価対象の分野に精通した二人の評価者が独立して評価した。すなわち、評価者 A が「内容の伝達レベル」と「重要技術用語の翻訳精度」を評価し、それと独立して評価者 B も同様の評価を実施した。そして「内容の伝達レベル」、「重要技術用語の翻訳精度」とも、評価者 A と B の平均値を各文の評価スコアとして採用した。

「内容の伝達レベル」および「重要技術用語の翻訳精度」の評価は、特許庁の「特許文献機械翻訳の品質評価手順」に記載の下記評価基準に則り実施した。

1.3.4.1 「内容の伝達レベル」の評価

「内容の伝達レベル」は、下記の 5 段階で評価した。

- | |
|--|
| 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%) |
|--|

1.3.4.2 「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」は、下記の 4 段階で評価した。

- | |
|--|
| 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。 |
|--|

なお、「重要技術用語の翻訳精度」の評価スコア A、B、C、D は、集計の際、A を 5 点、

B を 3 点、そして内容理解の役に立たない C と D はいずれも 0 点として数値化した。

1.3.4.3 評価者の基準合わせの方法

二次評価は、技術分野ごとに異なる評価者を起用し、また各文について二名の評価者による重複評価を行った。このため、各評価者間の評価基準をあらかじめ整合させて評価に臨む必要があった。

本評価においては、二次評価の実施に先立ち、全ての評価者に対し、以下の各手法にて評価基準の整合を行った。

①過去の評価結果を利用した実例集の作成・配布

「特許文献機械翻訳の品質評価手順」の「内容の伝達レベル」と「重要技術用語の翻訳精度」に基づく既存の評価結果を用いて、ここからそれぞれの観点における各評点に該当する実例集を作成し、あらかじめ全評価者に配布した。

②練習問題による事前の基準整合

同様に、既存の評価結果から適切な実例を採取し、これを練習問題として事前に全ての評価者に評価させた。その結果、担当者の定める正解評点とズレがある評価者に対しては、あらかじめ評価基準を調整し、全評価者の評価基準を整合させた。

③評価基準の説明

評価に先立ち、本評価の手順や評価基準、注意点についての説明会を開催し、原則として全評価者を参加させた。

1.3.5 二次評価結果（評価対象文単位）

二次評価では、全ての評価対象文に対し、「内容の伝達レベル」「重要技術用語の翻訳精度」とも二名の評価者にて重複評価を実施した。各文の評価スコアの算出にあたっては、下記の手順で複数の評価結果の平均値を取得し、採用した。

- ① 各文について二名の「内容の伝達レベル」の評価スコア（5～1 点）の平均を算出
- ② 同じく二名の「重要技術用語の翻訳精度」の評価スコアを数値換算（A→5 点、B→3 点、C&D→0 点）し、平均を算出
- ③ 前記①、②の平均値を合算して当該文の総合評価スコア（10～1 点）を算出

二次評価における個々の評価対象文の「内容の伝達レベル（5～1 の 5 段階評価）」及び「重要技術用語の翻訳精度（A～D の 4 段階評価）」の評価結果は、下記電子ファイルを参照されたい。

『01 二次評価（中 AU）文単位評価結果.xlsx』

なお上記ファイルには、各文について二名の評価者それぞれの評価結果を併記している。次ページの図 1-18 に、実際の評価作業にて納品された文単位の評価結果のサンプルを示す。

図 1-18：文単位評価結果サンプル（中 AU「ZDZ2-1」の 50 文より抜粋）

中AU	テーマ	原文	機械翻訳文	基準翻訳文	主要技術用語 (中)	主要技術用語 (MT)	主要技術用語 (日)	評価者1 (内容) 5~1	評価者1 (用語) A~	コメント(任意) 評価根拠等	評価者2 (内容) 5~1	評価者2 (用語) A~C	コメント(任意) 評価根拠等
ED12-1	5D182	本發明提供了一種結構簡單、調節方式簡便、支持作用比較強的伸縮音架。	本発明は構造が簡単であるを提供した、調節方式は簡便であり、デュアル比較的に強い伸縮音架スタンドを提供する。	本発明は構造が簡単で、調節方法も簡便で、伸縮効果の強い伸縮音架台を提供する。	伸縮音架	伸縮音架スタンド	伸縮音架台	2	A		2	A	
ED12-1	5D182	参照図1至圖3，本發明的馬林巴琴架包括音架兩側的两个支撐架1以及連接两个支撐架1的橫梁2。	図1〜図3を参照し、本発明のマリンバ琴架は両側面に設置する2個の支持枠1を含んで並びに2個の支持枠1の横梁2を接続する。	図1から図3に参照されたとおり、本発明のマリンバ台は両側に置かれた2台の支持スタンド1および2台の支持スタンド1を連接するビーム2を含む。	馬林巴琴架	マリンバ琴架	マリンバ台	2	C	カタカナ文じりの中国語まま	2	C	
ED12-1	5D182	本發明音架包括1-3上面固定好音架1-4(2-4)后并安装在音架箱室内。	本発明はバンド架スクロール軸1〜3上面で本発明のバンド架1〜4(2〜4)後に且つ本発明のリボンカセット内に取り付けることを好む。	本発明は軸1〜3上に本発明の紙1〜4(2〜4)を巻きつけてそして本発明の紙箱内に装着する。	本發明音架	本發明音架	本發明音架	2	C		1	C	
ED12-1	5D182	本發明能够适应更多的人群，使得多音鼓能够根据个人的习惯进行音量調節。	本発明は更に多い人々に適応することができ、多音にふくれている個人の習慣によって位置調節を行うことができる。	本発明は個人の好みに応じて多音タムの位置を調節できるため、より多くの人々に適合させることができる。	多音鼓	多音	タム	1	C	本発明は許容登録してください	1	C	
ED12-1	5D182	更优选地，每一琴架上均设有一对琴弦压板，所述琴弦压板通过弹性铰链与琴架固定铰接。	更に好ましく、1ごとの琴の音源板上で全部一方の琴の音源クリップを設け、前記琴の音源クリップは弾性ヒンジと琴の音源プレート固定によってつながる。	さらに好ましいのは、各楽器基に一方の楽器押さえ板を設け、前記楽器押さえ板を弾性ヒンジと楽器基に固定接続させることである。	琴弦压板	琴弦押さえ板	琴の音源クリップ	2	C		2	C	
ED12-1	5D182	本發明的有益效果是：通过支撐杆的设置，使得音架高度可調節，可滿足不同音量的需求。	本発明の利点は以下の通りである：支持ロッドの設置によって、音面の高さを調節することができる、異なる音量の需要を満たすことができる。	本発明の有益な効果は以下のとおりである：支持支柱の設置によって音面の高さを調節が可能となり、異なる音の要件を満たす。	音架	音架	音架	3	A		3	A	
ED12-1	5D182	本發明涉及一种乐器架，其中，该乐器架可容纳他、小提琴、大提琴和钢琴等。	本発明は楽器フレームに関し、そのうち、この楽器はギター、バイオリン、チェロとサクソフーンなどを収めることができる。	本発明は楽器スタンドに関し、ここにおいて、楽器はギター、バイオリン、チェロやサクソフーンである。	乐器架	楽器フレーム	楽器スタンド	3	B		2	C	
ED12-1	5D182	随着电子技术的发展，现在的钢琴也被赋予更多的功能，电子乐器也随着网络的发展而被更多的人使用。	電子技術の発展につれて、今のピアノはより多くの機能を付与されて、電子楽器はネットワークの発展によって多くの人が用いられることにつれて。	電子技術の発展に伴い、現在のピアノには様々な機能が付与されており、電子楽器もインターネットの発展と共に多くの人が使用されている。	电子乐器	電子の楽器	電子楽器	2	A		3	A	
ED12-1	5D478	随着电子技术的不断发展，以智能钢琴为代表的新一代智能电子乐器应运而生。	電子技術の絶えず発展することにつれて、インテリジェントな電子ピアノをはじめとして次世代インテリジェントな電子楽器は時速に広げられる。	電子技術のたゆまぬ発展に伴い、スマート電子ピアノに代表される次世代のスマート電子楽器が台頭し始めた。	智能电子乐器	インテリジェントな電子楽器	スマート電子楽器	4	A		4	A	
ED12-1	5D478	模数转换单元444，与所述译码单元442连接，用于将所述乐器音信号由模拟信号转换为数字信号；	A/Dの変換ユニット444は、前記スクリーニング部442に接続され、アナログ信号から前記楽器音信号をデジタル信号に変換することを用いる。	アナログデジタル変換ユニット444は、前記スクリーニングユニット442と接続され、前記楽器音信号をアナログ信号からデジタル信号に変換することを用いられる。	模数转换单元	A/Dの変換ユニット	アナログデジタル変換ユニット	4	B	専門用語だけに「の」は不要	4	B	
ED12-1	5D478	进一步地，所述音源播放板连接有音频输出插孔、耳机插孔和音量调节按钮。	さらに、前記音源再生板は音声出力ジャック、ヘッドフォンジャックと音量調整ボタンが接続されている。	さらに、前記音源再生ボードはオーディオ出力ジャック、イヤホンジャックと音量調節の回転ノブに接続される。	音频输出插孔	音声出力ジャック	オーディオ出力ジャック	4	A		4	A	
ED12-1	5D478	在实施例一的基础上，所述电热板3与电源2之间设有电阻值可调开关4。	実施例の一基礎上に、前記加熱板3と電源2の間で抵抗ポジションスイッチ4を設ける。	実施例1を基礎として、前記加熱板3と電源2の間に電気抵抗シフトポジションスイッチ4が置かれる。	档位开关	ポジションスイッチ	シフトポジションスイッチ	3	A		4	A	
ED12-1	5D478	进一步，所述演奏面包括演奏面上部和演奏面下部，并且演奏面上部的直径大于演奏面下部的直径。	さらに、前記ピストンカバーはピストンキャップ上部とピストンキャップ下部を含み、そしてピストンキャップ上部の直径はピストンカバー下部の直径より大きい。	さらに、前記ピストンキャップはピストンキャップ上部とピストンキャップ下部を含み、かつピストンキャップ上部の直径はピストンキャップ下部の直径より大きい。	演奏面	ピストンキャップ	ピストンキャップ	5	A		4	A	カバーとキャップの2種に訳されている

1.3.6 二次評価結果の集計（中 AU 単位）

上記 1.3.5 に示した評価対象文単位の評価結果を集計し、中 AU 単位での平均評価スコアを算出した。

算出は以下の手順で実施した。

[1. 文単位の平均値を算出]

- ① 各評価対象文について、二名の評価者の「内容の伝達レベル」の評価スコアの平均値を算出し、これを当該文の「内容の伝達レベル」の評価スコアとした。
- ② 各評価対象文について、二名の評価者の「重要技術用語の翻訳精度」の評価スコアを、A=5 点、B=3 点、C=0 点、D=0 点に換算のうえ平均値を算出し、これを当該文の「重要技術用語の翻訳精度」の評価スコアとした。

[2. 中 AU 単位の平均値を算出]

- ③ 各中 AU に属する 50 文について、①で算出した「内容の伝達レベル」の評価スコアの平均値を算出し、これを当該中 AU の「内容の伝達レベル」の評価スコアとした。
- ④ 各中 AU に属する 50 文について、②で算出した「重要技術用語の翻訳精度」の評価スコアの平均値を算出し、これを当該中 AU の「重要技術用語の翻訳精度」の評価スコアとした。
- ⑤ 各中 AU について、③で算出した「内容の伝達レベル」の評価スコアと、④で算出した「重要技術用語の翻訳精度」の評価スコアとを 1 : 1 の比率で合算し、これを当該中 AU の機械翻訳精度の総合評価スコア（10 点満点）とした。

二つの評価観点の重みづけは、⑤に記載のとおり、1 : 1 とした。本事業の成果物のうち、「中韓文献翻訳・検索システム」の翻訳精度向上のために即時に利用できるのが「中日対訳辞書データ」であることから、同等レベルの翻訳精度であれば、辞書データでは直接改善できない「エンジンのロジックに起因する誤訳」よりも、辞書データで改善できる「技術用語の誤訳」が生じている中 AU を優先的に選定すべきと考えた結果、特定の一語の評価である「重要技術用語の翻訳精度」に、「内容の伝達レベル」と同等の重みを与える（すなわち、1 : 1）ことが適当と判断した。

また、この重みづけは昨年度事業にて適用されたものと同一であるため、両者の評価スコアの直接対比できるという利点も有する。

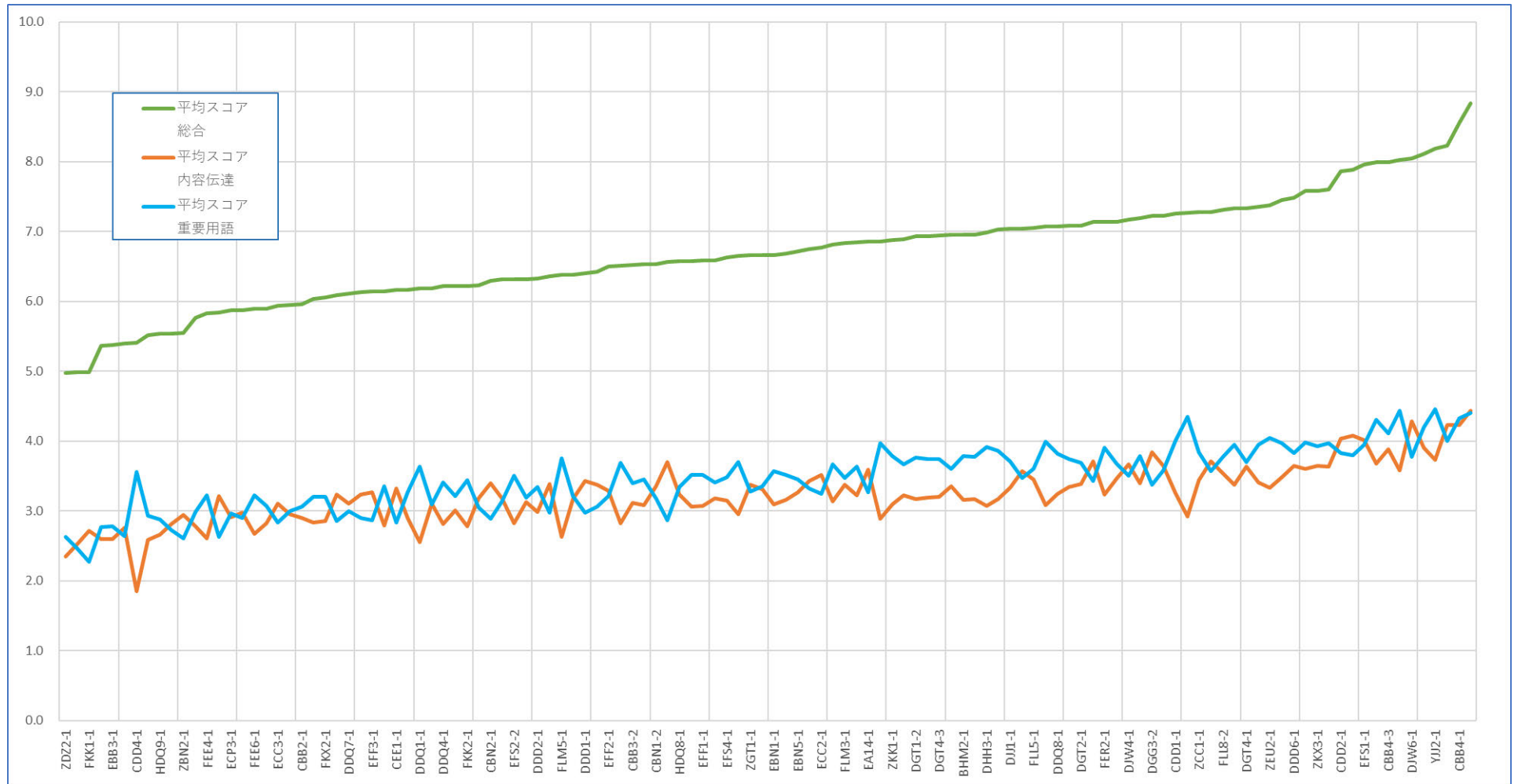
上記手順にて算出した中 AU ごとの総合評価スコアの低い順（品質が悪いと評価された順）に下位 30 位までの一覧表と、全中 AU を総合評価スコア順に並べたグラフを次ページの表 1-19、図 1-20 に示す。全中 AU（120 分類）の総合評価スコアは別添 1-1「二次評価

（中 AU）分類単位評価結果」を参照されたい。なお、本評価結果の分析は次の「1.3.7 二次評価結果の分析」の項にて述べる。

表 1-19：二次評価結果（中 AU 別平均評価スコア下位 30 位）

順位 (下位順)	中AU		平均スコア 総合	平均スコア 内容伝達	平均スコア 重要用語
1	ZDZ2-1	楽器	4.98	2.35	2.63
2	CER4-1	家具	4.99	2.52	2.47
2	FKK1-1	コネクタ	4.99	2.72	2.27
4	HDD1-1	固体廃棄物	5.37	2.60	2.77
5	EBB3-1	紙送り	5.38	2.60	2.78
6	ECC1-1	工作機械	5.4	2.76	2.64
7	CDD4-1	掘削	5.41	1.85	3.56
8	CBN5-1	パチンコ	5.52	2.59	2.93
9	HDQ9-1	混合	5.54	2.66	2.88
9	EFF5-1	噴霧・塗布	5.54	2.81	2.73
11	ZBN2-1	ケーブル処理	5.55	2.94	2.61
12	ECC5-1	携帯工具	5.77	2.78	2.99
13	FEE4-1	衛生・介護	5.83	2.61	3.22
14	CBB4-2	木材加工	5.84	3.21	2.63
15	ECP3-1	プレス加工	5.87	2.91	2.96
15	DJW2-1	ブレーキ	5.87	2.97	2.90
17	FEE6-1	医薬品の注入・保存・	5.89	2.67	3.22
17	EBN4-1	大型・ガス貯蔵容器	5.89	2.82	3.07
19	ECC3-1	マニプレータ	5.94	3.11	2.83
20	FER3-1	陳列棚	5.95	2.95	3.00
21	CBB2-1	収穫・脱穀	5.96	2.90	3.06
22	HCC3-2	動植物物質含有医薬	6.04	2.84	3.20
23	FKX2-1	スイッチ	6.06	2.86	3.20
24	CEE3-1	特定目的建築物	6.09	3.23	2.86
25	DDQ7-1	流体分配器	6.11	3.11	3.00
26	JHV5-1	化合物のその他の用途	6.13	3.23	2.90
27	EFF3-1	破碎・粉碎	6.14	3.27	2.87
27	DGG3-3	内燃機関本体	6.14	2.79	3.35
29	CEE1-1	建築構造・部材	6.16	3.32	2.84
30	FEE5-1	物理療法	6.17	2.90	3.27

図 1-20：二次評価結果（中 AU 別平均評価スコア順グラフ 全 120 分類）



1.3.7 二次評価結果の分析

上記 1.3.6 にて集計した中 AU 単位の人手評価スコアに基づき、二次評価結果の分析を実施した。

1.3.7.1 評価下位 10 分類の分析

まずは、二次評価の対象とした 120 の中 AU において、総合評価スコアが最も低かった 10 分類を下表 1-21 示す。

表 1-21：二次評価下位 10 分類

順位	中 AU		総合評価	内容伝達		重要用語	
			スコア	スコア	順位	スコア	順位
1	ZDZ2-1	楽器	4.98	2.35	2	2.63	4
2	CER4-1	家具	4.99	2.52	3	2.47	2
2	FKK1-1	コネクタ	4.99	2.73	12	2.27	1
4	HDD1-1	固体廃棄物	5.37	2.60	6	2.73	8
5	EBB3-1	紙送り	5.38	2.60	7	2.78	9
6	ECC1-1	工作機械	5.40	2.76	13	2.64	6
7	CDD4-1	掘削	5.41	1.85	1	3.56	66
8	CBN5-1	パチンコ	5.52	2.59	5	2.93	19
9	HDQ9-1	混合	5.54	2.66	10	2.88	15
9	EFF5-1	噴霧・塗布	5.54	2.81	17	2.73	7

上表のとおり、下位 10 分類の技術構成は、C で始まる分類と E で始まる分類が各 3 件、H が 2 件、Z と F が各 1 件となっている。ただし、この結果を見る際に留意すべきは、二次評価は全ての中 AU を対象としたのではなく、一次評価で粗ぶるいされた 120 の中 AU のみを対象としている点である。二次評価対象となった 120 分類は、下表 1-22 のとおりの構成となっている。

表 1-22：二次評価対象中 AU の構成

	A	B	C	D	E	F	G	H	J	R	Y	Z	計
二次	1	3	23	32	25	20	0	4	1	0	2	9	120
全件	15	36	28	49	35	34	21	22	25	3	26	32	326

上表のとおり、そもそも C、D、E、F は他に比べて一次評価での評価値が低かったため 120 分類における絶対数自体が多くなっており、それぞれほぼ 20%を占めている。このため、下位 10 位においても C や E が他よりも多数となることは当然といえる。むしろ、最も

分類数の多い D 分類が下位 10 分類に 1 件も入っていない点の特異である。このことから、自動評価においては D 分類がやや実際よりも低めの評価が下される傾向にある可能性がある。

これを踏まえ、二次評価の下位 10 分類が、自動評価指標（BLEU、RIBES）を用いた一次評価にてどのような評価であったか、および、一次評価の下位 10 分類が、二次評価でどのような評価であったかを分析した。下表 1-23 及び 1-24 にその結果を記す。

表 1-23：二次評価下位 10 分類の一次評価における評価順位

順位	中 AU		一次評価順位（※）		
			総合	BLEU	RIBES
1	ZDZ2-1	楽器	70	69	68
2	CER4-1	家具	7	5	9
2	FKK1-1	コネクタ	43	57	24
4	HDD1-1	固体廃棄物	215	191	242
5	EBB3-1	紙送り	6	9	3
6	ECC1-1	工作機械	56	50	67
7	CDD4-1	掘削	54	65	42
8	CBN5-1	パチンコ	13	13	16
9	HDQ9-1	混合	40	30	49
9	EFF5-1	噴霧・塗布	39	35	41
平均順位			54.3 位		

※一次評価順位も全て下位順（品質下位から 1 位、2 位…）としている

表 1-24 一次評価下位 10 分類の二次評価における評価順位

順位	中 AU		二次評価順位		
			総合	内容伝達	重要用語
1	EBN2-1	容器の蓋	64	89	45
2	CBB3-2	水産	49	45	51
3	EFS1-1	自動組立	111	114	104
4	ECC5-1	携帯工具	12	14	23
5	EBN1-1	包装・紙製品製造	61	40	67
6	EBB3-1	紙送り	5	7	9
7	CER4-1	家具	2	3	2
8	EBN2-2	容器	84	39	109

9	EFF3-1	破碎・粉碎	27	69	13
10	DJW4-1	固着	91	105	62
平均順位			50.6 位		

上の2つの比較表にあるとおり、一次評価と二次評価の両方で下位10位に入っていたのは2分類(CER4-1、EBB3-1)にとどまり、双方の結果に有意な関連性は見られない。特に、一次評価においては下位10分類中の7分類をE分類(加工・製造技術関連)が占めるという顕著な傾向が見られたが、二次評価ではそのような傾向は無くなった。このことから、自動評価の場合、E分類において実際よりも低めの評価となる傾向があると推論される。

1.3.7.2 評価上位10分類の分析

続いて、同様に二次評価における総合評価スコアの上位10分類を下表1-25に示す。なお、本項では順位は評価の低い順としているため、二次評価対象120分類中で最も評価スコアの良かった中AUが120位となり、以下、119位、118位…となる。

表 1-25：二次評価上位10分類

順位	中 AU		総合評価	内容伝達		重要用語	
			スコア	スコア	順位	スコア	順位
120	DDD4-1	船舶	8.84	4.44	120	4.40	118
119	CBB4-1	栽培、捕獲・駆除	8.56	4.23	118	4.33	116
118	DJW7-1	緩衝・防振	8.23	4.23	117	4.00	111
117	YJJ2-1	移動無線通信システム	8.19	3.73	110	4.46	120
116	DJJ3-1	伝動装置の配置・操作	8.11	3.91	113	4.20	114
115	DJW6-1	シール・圧力容器	8.05	4.28	119	3.77	88
114	CER1-1	施工	8.02	3.58	97	4.44	119
113	CBB4-3	バイオ・その他	7.99	3.88	112	4.11	113
112	FLL4-1	乾燥	7.99	3.68	106	4.31	115
111	EFS1-1	自動組立	7.96	4.01	114	3.95	104

傾向を見ると、下位10位には1件も入っていなかったD分類が4件を占めているのが目立つ。これは先に推論した「D分類は自動評価でスコアが実際より低くなる傾向」をある程度裏づけているといえる。あとはC分類が3件、E、F、Y分類が1件ずつであり、おおむね、120分類において多勢を占める技術分野が妥当に混在している状況である。

なお、二次評価での評価上位は、上位とはいえ一次評価で粗ぶるいされた120分類の範

囲内でのものであるため、326 分類の全てを対象とした一次評価の上位との比較には意味がない。そこで、これら二次評価での評価上位 10 分類が、一次評価ではどのくらいのランクであったかを調査した。下表 1-26 にその結果を示す。

表 1-26：二次評価上位 10 分類の一次評価における評価順位

順位	中 AU		一次評価順位 (※)		
			総合	BLEU	RIBES
120	DDD4-1	船舶	65	58	74
119	CBB4-1	栽培、捕獲・駆除	170	132	214
118	DJW7-1	緩衝・防振	51	49	56
117	YJJ2-1	移動無線通信システム	123	221	181
116	DJJ3-1	伝動装置の配置・操作	32	39	26
115	DJW6-1	シール・圧力容器	52	52	54
114	CER1-1	施工	29	24	35
113	CBB4-3	バイオ・その他	284	278	295
112	FLL4-1	乾燥	83	73	98
111	EFS1-1	自動組立	3	4	4
平均順位			89.2 位		

※一次評価順位も全て下位順（品質下位から 1 位、2 位…）としている

一次評価順位の平均順位（下位順）は 89.2 位であり、前出の表 1-23「二次評価下位 10 分類の一次評価における評価順位」に示した二次評価下位 10 分類における一次評価順位の平均順位（54.3 位）と比較すると、一定の整合性は見られる。ただし、例えば 111 位の EFS1-1 は一次評価では下位順で 3 位であった等、個々の分類単位では一次と二次とで大幅な逆転現象も数多く見られ、両者の相関性は薄いと言わざるをえない。

1.3.7.3 補足的観点にて二次評価に加えた分類の分析

二次評価の対象とした分類のうち、一次評価での総合順位が 102 位以降の分類は、スコアの上では二次評価の対象から外されるべきところ、ファミリー率や出願件数等の補足的な評価観点から敢えて二次評価の対象とした分類である。以下、これに該当する全中 AU について、一次評価順位と二次評価順位との対比を行った。

[観点①：BLEU と RIBES の評価差が顕著であった 6 分類]

まずは、補足観点①「BLEU と RIBES の評価差が顕著であった分類」について検討する。この観点に該当する中 AU は下表に示す 6 分類である。これら 6 分類について、二次評価の順位と一次評価の順位（総合／BLEU／RIBES）とを対比した。なお、この観点については、もともと二次評価対象となっている中 AU は存在しなかった。

表 1-27 「BLEU と RIBES の評価順位差が 51 以上の分類」二次評価結果（観点①）

No.	中 AU		一次 BLEU	一次 RIBES	一次 総合	二次
1	CBB4-1	栽培、捕獲・駆除	195 C	113 B	170 C	119D
2	ZEU2-1	メモリ回路・信頼性	122 B	187 C	169 C	103 D
3	EA21-1	鋳造	181 C	120 B	175 C	68 C
4	ZKX3-1	増幅器	72 A	130 B	224 C	107 D
5	HDD1-1	固体廃棄物	136 B	85 B	215 C	4 A
6	ZCC1-1	ビデオ記録	117 B	168 C	186 C	97 D

これら 6 分類は、いわば「BLEU と RIBES とで評価が分かれたため、どちらが妥当かを別途検証する」ために二次評価の対象とした側面がある。ただし、その観点においても、BLEU のほうが二次評価順位に近いもの（No.2、6）と RIBES のほうが近いもの（No.1、3、4、5）とがあり、全体としてどちらがより妥当かを判断できるような結果とはならなかった。また、一次評価が全中 AU（326 分類）を対象としているのに対し、二次評価ではそのうちの 120 分類のみを対象としているため、両者を単純に順位比較するのが妥当であるかも微妙である。

そこで、一次、二次とも、対象分類を 4 分割し、それぞれの順位がどこに該当するかについてもチェックした。つまり、一次評価であれば、範囲 A：1～82 位、B：83～163 位、C：164～245 位、D：246～326 位の 4 分割、二次評価であれば、範囲 A：1～30 位、B：31～60 位、C：61 位～90 位、D：91 位～120 位となる。上表の各順位欄に併記された A～D が、その範囲を示している。

ただしこちらの結果も、BLEU がより近いものが 2 件 (No.1、3)、RIBES がより近いものが 3 件 (No.2、4、6)、同点が 1 件 (No.5) とバラバラであり、やはりどちらの評価がより妥当かを判断できる全体的な傾向は見られなかった。

以上の結果からは、今後の自動評価の利用の際には、可能であれば本事業と同様に、BLEU と RIBES を併用し、双方を同じ比率で平均したスコアを採用するのが無難であると考え。

[観点②～③：日本における審査件数／審査請求率上位にあたる分類]

補足観点②～③では、「日本における審査件数／審査請求率上位の中 AU」ということで、る下表 1-28 の計 14 分類について、二次評価に追加した。なお、表中グレーで表示した各分類は、一次評価の順位が低く、もともと二次評価対象となっていたものである。

表 1-28 二次評価にて優先的に対象とする中 AU 一覧 (観点②～③)

No.		中 AU		一次	二次
1	②日本における審査件数上位	CBN5-1	パチンコ	13	8
2		YJJ2-1	移動無線通信システム	204	117
3		ZEE2-1	MM I	181	104
4		ZGT1-1	電力系統	191	59
1	③日本における審査請求率上位	CBN5-1	パチンコ	13	8
2		CBN5-2	スロット	16	38
3		CDD6-1	トンネル	49	66
4		ECC2-1	工作機械の制御	183	65
5		EFS4-1	実装	25	57
6		YJJ2-1	移動無線通信システム	204	117
7		CBN2-1	ゲーム・玩具	96	37
8		ZKK1-1	符号変換	238	71
9		FEE3-1	処置具	76	92
10		CDD3-1	基礎工	48	73

上表のとおり、観点②、③とも、二次評価に追加した 5 分類は概ね二次評価の順位が 120 分類中 60 位以降と、比較的評価の高いグループに属しており、その一方で、もともと二次評価対象であった残りの分類は、上記 5 分類よりは評価が低い傾向にある。このため、大枠としては一次評価の結果を裏付ける形になっているといえる。これは、そもそも観点②、③自体、「品質が相対的に低くなるはず」という推測のもと選定されたものではなく、我が国にとって特に重要な技術分野であったり、自動評価スコアにやや不安な点が見られたために、念のため二次評価で再確認したものであることを考えると、順当な結果ともいえる。

〔観点④：日本語ファミリー率が1%以下の分類〕

一方、補足観点④「日本語ファミリー率が1%以下」という条件は、「日本語ファミリー率が低い分類はそれだけ言語資源が乏しいため、本来、機械翻訳品質は相対的に低くなるはず」という推定に基づいている点で、観点②、③とは異なる。観点④の結果を下表 1-29 に示す。

表 1-29 補足的観点から二次評価対象とした中 AU の評価結果（観点④）

	中 AU		一次	二次
④日本語ファミリー率が1%以下	CDD6-1	トンネル	49	66
	CBB4-3	バイオ・その他	284	113
	HCC3-2	動植物物質含有医薬	220	22
	CBB4-2	木材加工	252	14
	JHV5-1	化合物のその他の用途	298	26

上表のとおり、観点④に基づき追加した4分類のうちの3分類が二次評価で10～20位台（下位順）とかなり低い評価となっており、一次評価とは大幅に異なる結果となった。サンプル数が4分類とごく少数であるとはいえ、今回の結果には一定の傾向が見て取れ、「日本語ファミリー率が低い分類は他に比して機械翻訳品質が低くなる」という推論は相応の信憑性を有していると判断できる。またその一方で、人手による精密な評価である二次評価にて一次評価と大幅に異なる傾向が示されたことから、自動評価の精度は、たとえ大量のデータを使用したとしても完璧とは限らないと考えておくべきであり、今後同様の評価を実施する際にも、本事業で実施したような補足的な観点を適宜組み合わせるべきである。

〔観点⑤：昨年度評価にて品質下位となった分類〕

補足観点としてはもう 1 点、観点⑤「昨年度評価（中 AU 単位）にて品質下位となった分類」についても二次評価対象に加えた。ただし、対象とした 11 分類中 8 分類（表中グレー表示）については、もともと自動評価（本年度一次評価）でも品質下位であったため無条件で二次評価対象となっており、追加したのは 3 分類(DGT1-1、DHH3-1、DDD6-1)のみであった。

表 1-30 補足的観点から二次評価対象とした中 AU の評価結果（観点⑤）

	中 AU		一次	二次
⑤昨年度 一次評価 品質下位	DGT1-1	タービン	148	72
	ECP2-1	非金属の加工	14	50
	DHH3-1	交通制御	203	79
	CBN5-1	パチンコ	13	8
	ECP7-1	装飾技術	90	48
	DDD6-1	武器	106	105
	DJW4-1	固着	10	91
	FKK1-1	コネクタ	43	2
	CBN1-1	遊具	58	100
	EFF3-1	破碎・粉碎	9	27
	ZBN2-1	ケーブル処理	42	11

上表を見ると、追加で二次評価対象とした 3 分類はいずれも評価順位（下位順）が 70 位台以下となっており、一次評価の順位よりは順位が上がって（評価が下がって）いる。ただし、先述のとおり一次評価と二次評価では対象とした分類数が異なるため、これをもって「二次評価は一次評価よりも昨年度評価に近づく傾向を見せた」とは言い切れない。また、もともと二次評価対象としていた 8 分類に関しては、昨年度評価に近づく傾向を示すものと遠ざかる傾向を示すものが半々という状況であり、やはり明確な傾向は見られない。

ただし、今回の結果をもって、過去の評価結果については今後考慮不要と結論するのは早計であると考え。これまでの分析から、今回の自動評価結果は人手による二次評価結果とあまり相関性は高くなく、したがって自動評価の精度には限度があると言わざるを得ない。そうすると、特に昨年度評価結果のように人手評価による高い信憑性を有するものである場合は、今後も本事業と同様、過去の評価結果は補足的な観点として重視すべきであろう。

1.3.7.4 翻訳品質が低いテーマコードの分析① 重要技術用語の翻訳精度

本評価では、「内容の伝達レベル」と「重要技術用語の翻訳精度」の評価スコアを意図的に1:1の重みづけで合算して総合評価としている。このため、各文で「重要技術用語」として選定された語の翻訳精度は、それだけで総合評価の50%を占め、さらに当然ながら「内容の伝達レベル」の評価にも加味されるため、双方の観点でダブルカウントされることとなる。このため、本評価において「重要技術用語」の翻訳精度は、翻訳品質の優劣評価に非常に大きな影響力をもつ。

これを踏まえ、二次評価の対象とした120分類中、まずは「重要技術用語」の平均スコアの下位10分類について、その傾向を分析した。分析対象となった中AU及び各中AUの平均スコアを下表1-31に示す。

表 1-31 「重要技術用語の翻訳精度」下位 10 分類

中 AU		用語		内容		総合	
		スコア	順位	スコア	順位	スコア	順位
FKK1-1	コネクタ	2.27	1	2.72	12	4.99	2
CER4-1	家具	2.47	2	2.52	3	4.99	2
ZBN2-1	ケーブル処理	2.61	3	2.94	29	5.55	11
ZDZ2-1	楽器	2.63	4	2.35	2	4.98	1
CBB4-2	木材加工	2.63	5	3.21	60	5.84	14
ECC1-1	工作機械	2.64	6	2.76	13	5.4	6
EFF5-1	噴霧・塗布	2.73	7	2.81	17	5.54	9
HDD1-1	固体廃棄物	2.77	8	2.60	6	5.37	4
EBB3-1	紙送り	2.78	9	2.60	7	5.38	5
ECC3-1	マニプレータ	2.83	10	3.11	43	5.94	19
(参考)	120 分類平均	3.52	-	3.28	-	6.80	-

これら10分類は当然ながら総合評価も低く、上表に示したとおり、下位1位～6位を占めるほか、残りの各テーマも全て下位21位以内となっている。

これら 10 分類における「重要技術用語」の訳語（機械翻訳結果）を概観した印象として、以下が挙げられる。

①名詞と解釈されなかった重要技術用語を含む

本評価では、「重要技術用語」として原文文字数が 3～6 文字の語を選定している。その結果、選定される重要技術用語は、ほぼ確実に 2 語以上からなる複合名詞となる。

「重要技術用語の翻訳精度」の下位 10 分類の大半では、こうして重要技術用語として選定された複合名詞が、正しく複合名詞と解釈されなかったケースが複数件発生していた。以下、そのケースの実例をいくつか挙げる。

表 1-32：重要技術用語が複合名詞と解釈されなかった実例

中 AU	中国語	基準翻訳	機械翻訳	内容
FKK1-1	隔爆箱体	防爆箱本体	爆発することを隔てて箱体	2
CER4-1	护饰板	保護裝飾板	飾り板…かばう	1
ZBN2-1	导线展放	電線架設	リード線は施工	1
ZDZ2-1	吉他品丝	ギターフレット	ギターを～絹糸を品定めする	1
CBB4-2	炭化处理窑	炭化处理炉	かまどを処理	2.5
ECC1-1	蒙皮拉伸钣金	外板ストレッチ	気囊延伸した板金	2
EFF5-1	乙炔清吹设备	アセチレンパージ装置	アセチレンは清く装置	2
HDD1-1	清洗分类	洗浄分別	分類することを洗浄し	1
EBB3-1	履带输送装置	クローラ搬送装置	クローラはデバイス…送って	2
ECC3-1	近接开关	近接スイッチ	近くスイッチ	1

このように、複合語が複合名詞と解釈されず、単語レベルで個別に異なる品詞に解釈されてしまうような場合、その悪影響はその複合名詞のみにとどまらず、文全体の翻訳精度を著しく損なうことは言うまでもない。一例として、上表より ZDZ2-1「吉他品丝／ギターフレット」を含む文について実例を示す。

原 文	本发明所要解决的问题是现有技术中存在的不足之处，提供一种可轻松、方便地调节高度的 吉他品丝 。
基準翻訳文	本発明で解決しようとする問題は従来の技術に不足しているもので、手軽かつ簡単に高さを調節できる ギターフレット を提供することである。
機械翻訳文	本発明解決しようとした問題は従来技術中存在していた不足点であり、提供して気楽、便宜のように高度な ギター を調節して 絹糸を品定めする 。

上の実例に示すように、「吉他品糸」を「ギターフレット」という複合名詞ではなく、「ギター」「絹糸」という単名詞と「品定めする」という動詞に解釈してしまったことで、この語を含む句の全体の構文解釈に悪影響を及ぼし、「(ギターフレットを) 提供する」という骨格自体が崩れてしまっている。

こうした悪影響の結果、上記の 10 例に関しては、「重要技術用語の翻訳精度」は当然 C 評価 (0 点) であるが、これらの語を含む文全体の「内容の伝達レベル」の評価も、大半が 1 点～2 点と非常に低くなっている。つまり、このような用語が重要技術用語として選ばれた場合、二つの観点 (「重要技術用語の翻訳精度」と「内容の伝達レベル」) において厳しく減点されるため、こうしたケースが複数件発生したようなテーマは、おのずと評価順位が下位に偏ってくるものと考えられる。

なお、評価対象文には技術用語が複数個含まれることも多く、また重要技術用語の選定は恣意的に「機械翻訳の難しそうな」用語を選んだわけではないため、このように複合名詞と解釈されない用語が、評価対象文中に含まれるものの、重要技術用語としては選定されていないというケースも当然想定される。ただしこうした場合も、「内容の伝達レベル」での大幅な減点は免れないため、10 文中にこのような用語を多く含むテーマは、自然と評価下位に集まることとなる。

このような「複合名詞が個々の単語としてバラバラに解釈される」というタイプの誤訳は、同一の語が複数の品詞となり、各語の品詞を語順や文脈から判断する必要があるという中国語の性質を考えると、現状の機械翻訳ソフトに単語レベルで完璧な訳し分けをさせることは極めて困難と思われる。例えば先に例として挙げた「清洗分类」についても、単に「清洗」と「分类」をユーザ辞書に個別に「洗浄」「分別」と名詞登録したとしても、当然ながら他の品詞での辞書登録も併存しているため、そのうちのどの品詞を採用するかは、最終的にはエンジンによる構文解析に委ねられることとなる。したがって、エンジンが構文解析を正確に行わない限り、こうした単語レベルの辞書登録では、誤訳の解消にはつながらないことが多い。そもそも「清洗」や「分类」といった一般的な名詞はもともとシステムの基本辞書に登録されていると考えられ、これをユーザ辞書に改めて登録したとしても状況は改善しないと考えるのが妥当である。

このような状況において有効なのが、複合名詞でのユーザ辞書登録である。これはエンジンの仕様にもよるため一概には言えないが、一般的なルールベース方式のエンジンの場合、上例であれば「清洗分类」という単位で名詞登録をすることで、エンジンの構文解析の際に最長一致の原則が適用され、「清洗分类」という単位で名詞解釈され易くなる。特

に「清洗分類」などは「分別を清浄する」という意味で使われることは考えられず、積極的に複合名詞で登録すべきである。

本事業では、二次評価の結果を受けて、相対的に翻訳品質が低いと見なされた中 AU を対象に、大量の和文抄録を作成し、これをソースとして 10 万語の機械翻訳用辞書を作成することとなっているが、その際には複合名詞を中心に収集することで、文全体への悪影響の大きい「複合名詞がバラバラに訳される」タイプの誤訳を効率的に減少させるべきである。

②重要技術用語の単純な誤訳

重要技術用語を複合名詞（正確には、名詞の連鎖）と正しく解釈している（つまり、品詞自体は正しく解析できている）ものの、訳語が不適切なため C 評価というケースも、当然ながら下位テーマには多数存在している。以下、その一例を示す。このパターンの誤訳の場合、少なくともこれが直接の原因となって文全体の構文解析を崩すことはないため、理論上は、「内容の伝達レベル」を含めた総合評価への悪影響は上記①のパターンよりは小さいといえる。

表 1-33：重要技術用語が複合名詞のまま誤訳された実例

中 AU	中国語	基準翻訳	機械翻訳	内容
FKK1-1	受力現象	ストレス現象	受力現象	1.5
CER4-1	機械臂	機械式アーム	機械ブーム	1
ZBN2-1	引流線	ジャンパー線	ドレナージ線	1
ZDZ2-1	鋼絲弦	ピアノ線	鋼線の弦	1.5
CBB4-2	四角木	角材	四隅の木	2.5
ECC1-1	保険杆	バンパー	保険ロッド	2.5
EFF5-1	花洒装置	シャワーヘッド	スプリングラ装置	2
HDD1-1	有色金属	非鉄金属	有色金属	1
EBB3-1	线缆轴盘	ケーブルリール	ケーブル軸盤	1
ECC3-1	仿真控制	シミュレーション制御	エミュレーションコントローラ	2.5

ただし、文の主題となる重要技術用語が大きく誤訳されると、その時点で（文法的に成立している文でも）内容の理解は難しくなることも多く、結果的には「内容伝達レベル」が①のパターンと同等程度に低評価となる可能性は高い。実際のところ、「重要技術用語のみが誤訳でそれ以外の部分は正確」というケースはほとんど存在せず、全体的には「重要技術用語」が低評価の文は「内容伝達レベル」も低い評価に止まる傾向にあった。

こうした誤訳も、その大半は、品詞こそ正しいとはいえ、複合名詞を個々の名詞としてバラバラに翻訳した結果によるものといえる。たとえば上表に ZDZ2-1 の例として挙げた「钢丝弦」などは、「钢丝」→「鋼線」、「弦」→「弦」という単語レベルでは適訳となるケースも多いが、両者を組み合わせた複合語「钢丝弦」の場合は「鋼線の弦」では意味をなさず、複合名詞として「ピアノ線」と訳さなければならない。したがって、このパターンの誤訳に関しても、複合名詞の辞書登録が最善の解決策といえる。

③重要技術用語が未知語であるケースは少ない

機械翻訳において、未知語は翻訳精度を大きく損なう主要な要因の一つであり、これは中日翻訳においても原則として当てはまる。ただし中日翻訳の場合、漢字から漢字への翻訳が大半であるため、未知語と誤訳との境界が曖昧で区別しにくいこと（例えば上記②に示した「受力現象」⇒「受力現象」や「有色金属」⇒「有色金属」などは、未知語なのか誤訳なのか微妙）、また、同じ漢字なので未知語であっても意味がかなり把握できる場合もあるなど、未知語のみに限定した悪影響を抽出することはかなり困難である。

ただし本調査の場合、二次評価結果にて評価者が「未知語（D 評価）」と判断した重要技術用語は 6,000 語中 18 語とごく少ない。これには上記のように未知語なのか誤訳なのか判断が難しいためという理由もあるが、それを踏まえても「中韓文献翻訳・検索システム」においては未知語の発生率はかなり低いといえる。このため、評価結果全体への影響も局所的なものにとどまっていると見なせる。

④重要技術用語の語長と評価値の関係

本評価では、重要技術用語の語長の範囲を 3 文字～6 文字としている。ただし、その範囲で何文字の用語を選定するかは任意であるため、中 AU ごとの平均語長は一定とはいえない。このため、特定の中 AU に語長の長い用語が集中し、翻訳の難易度が不当に高くなってしまっているという可能性もある。

これを検証するため、全中 AU における重要技術用語の平均語長と、スコア下位 10 位における平均語長、スコア上位 10 位における平均語長の比較を行った。

本評価では全中 AU で 6,000 語の重要技術用語が選定されている。まずはこの 6,000 語の語長分布について下表 1-34 に示す。

表 1-34：重要技術用語 6,000 語の語長分布

語長	3 文字	4 文字	5 文字	6 文字	7 文字	平均語長
用語数	2,329	2,325	837	508	1	3.92 文字

これに対し、重要技術用語のスコア下位 10 位の中 AU と上位 10 位の中 AU それぞれの平均語長について下表に示す。

表 1-35：重要技術用語スコア下位 10 位・上位 10 位の中 AU 別平均語長

	スコア下位 10 位			スコア上位 10 位		
	中 AU	スコア	平均語長	中 AU	スコア	平均語長
1	FKK1-1	2.27	4.16 文字	YJJ2-1	4.46	4.02 文字
2	CER4-1	2.47	3.72 文字	CER1-1	4.44	3.58 文字
3	ZBN2-1	2.61	4.02 文字	CBB4-1	4.40	3.84 文字
4	ZDZ2-1	2.63	3.86 文字	DDD4-1	4.35	3.96 文字
5	CBB4-2	2.63	3.64 文字	EBN3-1	4.33	3.64 文字
6	ECC1-1	2.64	4.10 文字	FLL4-1	4.31	3.98 文字
7	EFF5-1	2.73	4.08 文字	DJJ3-1	4.20	3.78 文字
8	HDD1-1	2.77	3.94 文字	CBB4-3	4.11	3.58 文字
9	EBB3-1	2.78	4.02 文字	ZEU2-1	4.05	4.12 文字
10	ECC3-1	2.83	4.18 文字	CDD1-1	4.03	3.58 文字
10 位までの平均			3.97 文字	10 位までの平均		3.80 文字

この結果を見ると、スコア下位 10 位は全体平均語長 3.92 文字よりもわずかに長い 3.97 文字、上位 10 位は若干短い 3.80 文字であったが、いずれも微差であり、また個々の中 AU を見ると上位下位ともかなりのバラつきがあることを考えると、スコア下位ほど語長が長くなるという傾向があるとは見なせず、したがって上位と下位とで語長という条件面での有利不利はほぼ存在していないといえる。また、仮に重要技術用語の平均語長の差異が偶然によるものでないとしても、重要技術用語は所定の範囲内での無作為抽出であることから、選定された用語の平均語長が短い中 AU は、そもそも使用される用語に短めのもので多い（すなわち、翻訳難易度が低いため翻訳精度が高い）技術分野であるとも考えられる。このため、今回の重要技術用語の評価値は、各中 AU の技術用語の翻訳精度を示すものとして妥当であると結論される。

1.3.7.5 翻訳品質が低いテーマコードの分析② 内容伝達レベル

次に、もう1つの評価観点である「内容の伝達レベル」についても、平均スコアの下位10分類についてその傾向を分析した。分析対象となった中 AU 及び各中 AU の平均スコアを下表 1-36 に示す。

表 1-36 「内容の伝達レベル」下位 10 分類

中 AU		内容		用語		総合	
		スコア	順位	スコア	順位	スコア	順位
CDD4-1	掘削	1.85	1	3.56	66	5.41	7
ZDZ2-1	楽器	2.35	2	2.63	4	4.98	1
CER4-1	家具	2.52	3	2.47	2	4.99	2
DDQ1-1	サスペンション	2.55	4	3.64	73	6.19	31
CBN5-1	パチンコ	2.59	5	2.93	19	5.52	8
HDD1-1	固体廃棄物	2.60	6	2.77	8	5.37	4
EBB3-1	紙送り	2.60	7	2.78	9	5.38	5
FEE4-1	衛生・介護	2.61	8	3.22	39	5.83	13
FLM5-1	製氷・冷蔵庫	2.63	9	3.75	85	6.38	43
HDQ9-1	混合	2.66	10	2.88	15	5.54	9
(参考)	120 分類平均	3.28	-	3.52	-	6.80	-

①文中の技術用語の誤訳

前項では、「重要技術用語のスコアが低い場合、その影響を受けて内容伝達レベルのスコアも自ずと低くなる」という現象について考察したが、その逆、すなわち「内容伝達レベルのスコアが低い場合に重要技術用語のスコアも確実に低くなる」という現象も同様に存在するかとなると、こちらは必ずしもそのような関係性はないと考える。たとえば以下の実例のように、重要技術用語は A 評価（最高）だが内容伝達レベルのスコアは 1（最低）という結果となっているものも実際少なからず（6,000 文中 85 文）存在している。

原 文	PLC 可編程逻辑控制器，用于根据所述 CCD 图像控制器发送的所述偏差值计算出偏差的补正值；
基準翻訳文	P L C プログラマブルロジックコントローラは、前記 C C D イメージコントローラが送信した前記偏差値に基づいて、偏差の補正値を計算するのに用いられる。

機械翻訳文	CCDイメージコントローラは、基板上P I の塗布の実際位置を測量することをういて、且つ前記実際位置とデフォルト位置で比計算に対することを行い、ズレ値を得る；
-------	---

※赤字が重要技術用語

上の事例では、重要技術用語として選定された「图像控制器／イメージコントローラ」自体は正確に翻訳されているが、各要素の関係性がきわめて不正確であり、また「プログラマブルロジックコントローラ」という要素自体がバラバラに訳されて判別困難であるなどの理由で内容の伝達レベルは最低評価となっている。

このように、内容伝達レベルに影響する要素はさまざまであるが、本調査の場合、各テーマとも同じ翻訳エンジンを使用しており、翻訳のロジック自体は共通しているため、各分類とも同等程度の難易度の文がサンプリングされているという前提に立てば、それぞれの評価の差は、主にそこに含まれる技術用語が正しく解析・翻訳されているかの差と考えることもできる。上記 10 分類における「重要技術用語」の平均評価スコアは 3.06 であり、120 分類全体の平均スコア 3.52 よりも有意に低いことから、この考えが裏付けられる。さらに、技術用語は「重要技術用語」として選ばれた一語以外にも文中に複数存在していることが大半であり、例えば上例の「プログラマブル～」のように、こうした「重要技術用語以外の技術用語」が誤訳されている場合には、「内容の伝達レベル」に低い評価が下されることとなる。したがって、「内容の伝達レベル」の観点においても、技術用語の翻訳精度は非常に大きなウェイトを占めているといえる。

もちろん、技術用語が原因ではなく、その分野特有の文体の癖などによって翻訳精度が低くなっているというケースも当然考えられる。ただし、文体や文構造が原因の誤訳は、ルールベース翻訳の場合、原則としてエンジンのロジックを改善しない限り辞書登録では解決不可能である。このため本評価では、評価スコアの集計において重要技術用語の比重を相当に重くしており、その分、構文解析の比重は軽くなっている。よって、同じ程度の内容伝達レベルの文であれば、用語の訳質が原因であるもののほうが、それ以外の原因のものよりも下位か、少なくとも同等に扱われる仕組みとなっており、全体的に見れば、総合評価スコアは下位の中 AU は、文中に含まれる技術用語の翻訳精度が相対的に低いテーマであると見なせる。

②コロケーションの問題

中国語に限らず、文中の語の多くは二つ以上で結びつけて使われる。コロケーションとは、こうした特定の語同士の「よく使われる組み合わせ、自然なつながり」を指す。今回

調査対象とした中日機械翻訳において、「内容の伝達レベル」の評価に影響を与えたと考えられる不備の傾向の一つとして、この「コロケーションの問題」が指摘できる。

中国語では自然な単語同士の結びつきも、それを単語単位で個別に日本語に訳出してしまうと、日本語のコロケーションに該当しない不自然な語の結びつきとなり、評価者によっては「誤訳」と見なされたり、日本語として過度に不自然と判定されることとなる。つまり、個々の単語レベルでは使用頻度の高い訳語が選定されているものの、コロケーションの問題で低評価となってしまうケースもあるということである。

下表 1-37 に、この「コロケーションの問題」に該当する事例を挙げる。

表 1-37：コロケーションの問題に該当する事例と内容伝達スコア

中 AU	原文	基準翻訳	機械翻訳	スコア
CBN4-1	…轻松滑动。	…スムーズに摺動する。	…気楽に摺動する。	1.5
CBB4-2	…美观大方。	…美しい。	…美観はおっとりしている。	3.0
BCP1-1	…喷水汽…	…発生させた蒸気と…	…噴水する蒸気と…	2.0

残念ながら、こうしたコロケーションへの対応は、機械翻訳においては「一つの語を、共起する語により訳し分ける」仕組みが必要となり、ルールベース方式のエンジンの場合、主にエンジンのロジックでの対応となるため、ユーザ側では対応する手段が用意されていないことが多い。ただし、機械翻訳品質の改善のためにはいずれ対処すべき主要な課題の一つであり、ここで報告しておく。

③助詞（てにをは）の問題

機械翻訳において、「てにをは」、すなわち助詞の誤訳というのは非常に大きな課題である。ルールベース方式の場合、一般的には、文法規則に則って構文解析を行うため、助詞の用法は比較的正確（少なくとも一定のルールに則っている）といえるが、中国語の場合、格助詞という概念が原則存在せず、語順が文構造を解釈する際の重要な決め手となるという性質上、ルールベース方式を用いた今回の機械翻訳においても、低評価の原因となったと見なせる助詞の誤りが非常に目立った。

下表に、この「助詞（てにをは）の問題」に該当する事例をいくつか挙げる。

表 1-38：助詞（てにをは）の問題に該当する実例と内容伝達スコア

中 AU	原文／基準翻訳文／機械翻訳文		スコア
AGG4-1	原文	将氧化石墨烯部分还原，有利于提高膜的使用稳定性以及膜的通量和截流性能。	1.5
	基準	酸化グラフェンの部分還元は、フィルムの利用安定性及びフィルムの流束や遮断特性の向上に効果的である。	
	機械	酸化グラフェン部分的還元を、向上フィルムの使用安定性及び膜流束と堰き止めの性能に有利である。	
BCP2-2	原文	本发明涉及一种热转印色带，特别涉及一种用于金属材料的热转印色带。	2.5
	基準	本発明は熱転写リボンに関し、特に金属材料を用いる熱転写リボンに関する。	
	機械	本発明は熱転写インクリボンに関し、特に金属材に用いる熱転写インクリボンに関する。	
ZKX3-1	原文	该低温低噪声 CMOS 差分运算放大器的放大部分如图 1 所示，采用差分输入的一级折叠共源共栅结构。	1.5
	基準	該低温、低ノイズのCMOS差動演算増幅器の増幅部は図1に示すように、差動入力の一級のカスコード構造が採用されている。	
	機械	この低温低騒音CMOSの差動演算増幅器の拡大部の図1に示すように、差動入力を用いる一級はカスコード構造を折り畳む。	

特に2つめの例に関しては、誤訳箇所はほぼ赤字部分の助詞のみであるが、この一箇所の誤訳により文意は大きく変わってしまっているため、内容伝達レベルのスコアは2.5と大幅に下げられている。このように、助詞の誤訳はたとえ一箇所であっても文の内容把握において致命的なものとなりうる。

ただし、こうした「助詞の誤訳」についても、基本的にはエンジンのロジックに委ねられており、ユーザ側では原則対処不可能であるが、特に中日機械翻訳においては大きな課題となっている点であり、いずれ解決が必要である。

④四字熟語の誤訳

技術用語とはいえませんが、四字熟語（に代表されるイディオム）をそう解釈できず、原文の個々の漢字の意味で翻訳してしまい、結果として意味不明となっているケースも散見

された。こうした誤訳も、「内容の伝達レベル」の評価値に影響していると考えられる。
以下、実例を挙げる。

表 1-39：四字熟語の誤訳に該当する実例と内容伝達スコア

中 AU	原文	基準翻訳	機械翻訳	スコア
CBB3-2	老少皆宜	子供から老人まで	老若がすべてに適した	4.0
CBB4-1	春寒料峭	初春の肌寒さ	春寒でありうすら寒く	4.0
CBB3-2	修身养性	身を修め自己を磨き	身を修めて性格を養い	1.0

上表に挙げた各例は、何となく意味は把握できるせいか「内容伝達レベル」のスコアは意外と高いが、日本語文として非常に不自然であることは間違いなく、また四字熟語は特許明細書でもそれなりに使用されているようであるため、改善できるのであればするに越したことはない。前述のコロケーションと異なり、四字熟語は全体を一語として辞書登録できるものも多いので、ユーザ辞書により対応できる可能性が高い。ただし、品詞は単純に名詞登録できないものが大半と思われ、あとはエンジンがユーザ辞書登録において許容している品詞設定の範囲次第といえる。

⑤評価選定文の文長と評価値の関係

本評価では、評価対象文の文長範囲を 30 文字～50 文字と定め、かつ、極力 35～45 文字のコア範囲とするように文を選定し、各中 AU の評価対象文の文長という条件面での有利不利が生じないように努めた。

とはいえ、各中 AU において多少の差異が生じることは避けられず、この「中 AU ごとの平均文長の差異」が、内容伝達レベルの評価スコアに影響している可能性がある。これを検証するため、全中 AU における評価対象文の平均文長と、スコア下位 10 位における平均文長、スコア上位 10 位における平均文長の比較を行った。

本評価では全中 AU で 6,000 文の評価対象文が選定されている。まずはこの 6,000 文の文長分布について下表 1-40 に示す。

表 1-40：評価対象文 6,000 文の文長分布

文長	30～34	35～40	41～45	46～50	平均文長
評価文数	993 文	2,460 文	1,671 文	876 文	39.7 文字

上表のとおり、全 6,000 文の平均文長は 39.7 文字と、思惑どおりコア範囲の中央付近にあたる値となっている。これに対し、重要技術用語のスコア下位 10 位の中 AU と上位 10 位の中 AU それぞれの平均文長がどうなっていたかについて、下表 1-41 に示した。

表 1-41：内容伝達レベルのスコア下位 10 位・上位 10 位の中 AU 別平均語長

	スコア下位 10 位			スコア上位 10 位		
	中 AU	スコア	平均文長	中 AU	スコア	平均文長
1	CDD4-1	1.85	40.1 文字	DDD4-1	4.44	39.3 文字
2	ZDZ2-1	2.35	39.5 文字	DJW6-1	4.28	39.0 文字
3	CER4-1	2.52	38.8 文字	CBB4-1	4.23	39.9 文字
4	DDQ1-1	2.55	40.0 文字	DJW7-1	4.23	38.4 文字
5	CBN5-1	2.59	39.3 文字	DJJ4-1	4.08	40.0 文字
6	HDD1-1	2.60	37.4 文字	CDD2-1	4.03	38.7 文字
7	EBB3-1	2.60	39.2 文字	EFS1-1	4.01	39.1 文字
8	FEE4-1	2.61	39.7 文字	DJJ3-1	3.91	40.9 文字
9	FLM5-1	2.63	39.8 文字	CBB4-3	3.88	41.1 文字
10	HDQ9-1	2.66	38.0 文字	DGG3-2	3.84	37.5 文字
10 位までの平均			39.2 文字	10 位までの平均		

上表に示されたとおり、スコア下位 10 位（平均 39.2 文字）、上位 10 位（同 39.4 文字）とも、全体の平均文長である 39.7 文字と近似であり、評価対象文の文長という条件面においても、各中 AU 間に有利不利は存在しなかったといえる。このため、二次評価における各中 AU の評価値の差異は純粋に各分類間の翻訳精度の差異と考えてよいであろう。

1.3.7.6 技術分野ごとの傾向の分析

二次評価の対象とした中 AU は、それぞれが 2B～5Z に区分された「技術単位」に属している。さらに、この「技術単位」は、上 1 ケタ（2～5）で集約することで、特許庁内審査部門ごとの 4 つの「担当技術」に区分することができる。

二次評価の時点では、既に中 AU の粗ぶりがされている状況ではあるが、補足的な分析として、上記 4 つの「担当技術」別の評価結果について集計を実施した。下表に、その結果を示す。

表 1-42: 「担当技術単位（2～5）」別の中 AU 構成および平均評価スコア（120 分類）

担当技術	(参考)全中 AU		二次対象中 AU		平均順位	総合	内容	用語
2	79	24.2%	27	22.5%	57 位	6.65	3.21	3.44
3	116	35.6%	78	65.0%	61 位	6.72	3.24	3.48
4	67	20.6%	4	3.3%	30 位	6.10	2.97	3.13
5	64	19.6%	11	9.2%	74 位	6.88	3.35	3.53
合計	326	100%	120	100%	60.5 位	6.70	3.23	3.46

上記 120 分類における構成比では担当技術 3 が 65%と過半数を占めている。担当技術 3 の全中 AU（326 分類）における比率は 35.6%なので、一次評価においては他の担当技術よりも品質が顕著に低いと見なされたといえる。逆に担当技術 4 及び 5 は、全中 AU に占める比率よりも二次評価対象中 AU に占める比率を大きく減じており、一次評価において品質が良好と判定されたことがわかる。

ただし、この構成比はあくまで一次評価（自動評価）に基づいた結果であり、これをもって担当技術 3 が他の担当技術よりも品質が悪く、担当技術 4 と 5 が品質が良いとは言いきれない。実際、二次評価での平均順位や平均評価スコアを見ると、担当技術 3 は 4 技術中で 2 番目に評価が高くなっている。

二次評価においては、担当技術 2、3、5 はいずれも全体の平均とほぼ同じ範囲に属しており、今回の順位をもって各技術分野の優劣とするのは危険である。また、二次評価では唯一スコアが顕著に低い担当技術 4 に関しても、サンプルとなった中 AU 自体が 67 分類中の 4 分類と非常に少なく、これをもって担当技術 4 が最も品質の悪い技術分野と判断することはできない。

したがって今回の分析からは、担当技術という大きな括りでは、他よりも特段に品質が悪いと見なすべき担当技術は確認されなかったという結論となる。

1.4 和文抄録作成対象案件の選定

二次評価の集計結果に基づき、評価対象とした 120 の中 AU から翻訳精度が相対的に低い中 AU を特定し、この分類に属する中国特許文献を和文抄録作成対象として選定した。以下、選定の詳細について記す。

1.4.1 テーマコード単位の選定

本事業の二次評価は中 AU 単位で実施しており、和文抄録作成対象も中 AU 単位で評価が下位のものから順に選定することとなる。

ただしその一方で、各中 AU は下層分類としてテーマコードを有しており、前年度事業では和文抄録の作成対象分類はこのテーマコード単位で選定している。このため、本事業で和文抄録作成対象とする中 AU においても、そこに属するテーマコードの一部に対し前年度事業で和文抄録を作成済み、というケースが想定される。

このようなケースにおいて、当該中 AU 全体を「前年度事業で和文抄録作成済み：和文抄録作成不要」とみなすことは不適であり、テーマコードごとの要否判断が必要となる。このため、以降に説明する和文抄録作成対象案件の選定作業は、テーマコード単位で実施した。ただし、翻訳品質の順位付けはあくまで中 AU 単位であるため、それぞれの中 AU において、その下層の全てのテーマコード間に優先順位は設けず、全てのテーマを同等に取り扱っている。

1.4.2 和文抄録作成可能案件の特定

和文抄録作成対象案件の選定にあたり、まずは、中 AU を二次評価下位順に並べたうえで、それぞれを下層テーマコードに細分化し、各テーマコードについて、和文抄録の作成対象とすることができる案件、すなわち、「平成 15～27 年発行の中国公開特許公報（A 公報）に対応する中国特許公報（B 公報）」に該当する案件をカウントした。なお、カウントの際は、前年度事業ですでに和文抄録を作成済みの案件は除外している。

1.4.3 和文抄録作成対象案件 8 万件の選定

和文抄録作成対象案件は、原則として、中 AU の下位順に、その中 AU に属する全てのテーマコードの全ての作成可能案件（1.4.2 でカウントされた案件）を選定していき、合計 8 万件とした。ただし、選定の際には以下の調整を行った。

- ① 特定のテーマコードへの案件の集中を避けるため、各テーマコードについて、案件数の上限値を 2,000 件と定め、これを超えるテーマについては、登録日の新しい順に 2,000 件を選定した。例えば、中 AU「HDD1-1」（第 4 位）の下層テーマコードである 4D004 は、和文抄録作成可能案件は 4,717 件存在したが、上限値 2,000 件を超えているため、登録日の新しい順に 2,000 件のみを選定した。
- ② 前年度に和文抄録対象となったテーマコードについては、前年度作成済み件数との合算で上限を 2,000 件とした。例えば、中 AU「CEE1-1」（第 29 位）の下層テーマコードである 2E121 では、今回カウントされた和文抄録作成可能案件は 446 件であったが、前年度事業にて既に 1,621 件の和文抄録を作成済みにつき、上限値 2,000 件に抵触せぬよう、登録日の新しい順に 379 件のみを選定した。

上記①、②の基準で案件を選定していくと、第 48 位の中 AU「ECP7-1」までは、全ての下層テーマコードをカバーすることができる。この時点での累計件数は 79,624 件であり、本事業での和文抄録作成件数 80,000 件との差分は 376 件である。この 376 件は、第 49 位の中 AU「CBB3-2」から以下の手法で選定することとした。

- ③ CBB3-2 の各下層テーマコードについて、上記①、②を踏まえて本事業で和文抄録作成対象とできる上限件数をカウントした（なお CBB3-2 については、昨年度和抄作成対象となったテーマコードも、本年度和抄作成可能件数が 2,000 件を超えるテーマコードも含まれていないため、上記①、②に関しては特段の配慮は不要であった）。

CBB3-2 に属する各テーマコードの本年度の作成可能件数が、次表 1-43 の「作成可能件数」となる。さらに、この作成上限件数の総計（504 件）に対する各テーマコードの構成率（表中「③の構成比」）を求め、CBB3-2 全体としての作成件数である 376 件をこの比率に基づきテーマコードごとに割り振った件数が、同表の「作成件数」となる。

表 1-43:中 AU「CBB3-2」の下層テーマコード別 和抄作成件数

テーマ	作成可能件数	CBB3-2 内構成比	作成件数
2B019	26 件	5.2%	19 件
2B105	60 件	11.9%	45 件
2B106	34 件	6.7%	25 件
2B108	113 件	22.4%	84 件
2B109	171 件	33.9%	128 件
2B307	100 件	19.8%	75 件
合計	504 件	100%	376 件

表中のテーマコード 2B019 を例に採ると、作成可能件数は 26 件なので、CBB3-2 全体の作成可能件数 504 件に占める割合は約 5.2%となる。この比率を、作成件数 376 件に反映した件数（376 件×5.2%＝19 件）を、2B019 の作成件数とした。

CBB3-2 については、この方法で、各下層テーマコードの作成件数を決定した。

このようにして選定した、本事業の和文抄録作成対象 8 万件の中 AU／テーマコード別の件数を、別添 1-9「中 AU／テーマコード別 和文抄録作成件数一覧」に示す。なお、選定された 8 万件の詳細は電子ファイル『02 和文抄録作成対象 8 万件.xlsx』を参照されたい。

1.5 新旧の機械翻訳文の比較評価の実施

本事業の機械翻訳品質評価では、ここまで述べた一次評価、二次評価に加え、前年度の品質評価の結果、相対的に翻訳精度が低いと判定され、和文抄録の作成対象とされた 174 のテーマコードについて、各テーマに属する文を合計 500 文選定し、当時の機械翻訳システムによる翻訳結果と、最新の機械翻訳システムによる翻訳結果とをそれぞれ人手評価し、その品質の変化を分析する「新旧機械翻訳文の比較評価」についても実施した。

本項にて、その概要について説明する。

1.5.1 新旧機械翻訳文の比較評価の概要

前年度事業では、当時の品質評価結果に基づき、特許庁の「中韓文献翻訳・検索システム」において相対的に翻訳精度が低いテーマコードを特定し、それらのテーマに属する中国登録特許文献 8 万件について和文抄録を作成した。さらに、この和文抄録等をソースとして約 10 万語の中日機械翻訳辞書データを作成し、翻訳精度向上のための言語資源の充実を図った。

本評価の主旨は、当時の「中韓文献翻訳・検索システム」による機械翻訳結果と、同システムの機械翻訳エンジンに上記機械翻訳辞書データ（以下、「前年度辞書データ」という。）等を反映した最新の機械翻訳システムによる機械翻訳結果とを比較評価することで、双方の機械翻訳結果にどの程度の変化が生じており、それにより翻訳精度がどの程度改善／悪化しているかを分析することにある。

したがって、評価対象とするテーマコードは、前年度事業にて和抄作成対象とした 174 テーマとなる。評価対象とする文は、全テーマ合計で 500 文とした。

評価対象テーマコードの一覧は別添 1-10「新旧機械翻訳文比較評価対象テーマコード一覧」を参照されたい。

1.5.2 評価対象文献の選定

評価対象文 500 文は、評価対象 174 テーマに属する文献 500 件からそれぞれ 1 文ずつ抽出することとした。

抽出にあたり、まずは二次評価と同様、平成 27 年発行の中国公開特許公報から、「中国公開特許公報の機械付与テーマコードリスト（平成 15～27 年発行の中国公開特許公報に対しテーマコードを機械的に付与したテーマコードリスト）」に基づき、各文献の属するテーマコードを特定したうえで、評価対象となる 174 テーマそれぞれに属する文献を 3 件ずつ、計 522 件（174 テーマ×3 案件）選定し、評価対象文献集合とした。

1.5.3 評価対象文候補の選定

評価対象文の候補として、まずは上記 1.5.2 にて選定した 522 文献(174 テーマ×3 文献)を対象に、1 文献あたり 1 文ずつ、評価対象文を選定した。なお、各文献における評価対象文の選定基準は、二次評価と同様、以下の基準とした。

- A. 対象文の文長を一定範囲（30～50 文字）に揃える
- B. 対象文の抽出箇所を「発明の詳細な説明」に揃える
- C. 難易度を左右する特殊な表現を含む文は除外する
- D. 所定文字数（3～6 文字）の重要技術用語候補を含む文のみを対象とする

各基準の詳細は 1.3.2 を参照されたい。

1.5.4 評価対象文予備候補の準備

本評価の主旨は、新旧の機械翻訳結果の比較評価にある。特に、新旧の機械翻訳結果の精度の違いを分析するためには、新旧で機械翻訳内容の異なる文であることが重要である。

一方、評価対象文は、種々の条件に基づき選定しているが、その文の機械翻訳結果が新旧で変化するか否かについては事前に予見することができず、この点に関しては無作為抽出となる。これは、一定量のサンプル中での変化率を把握するにあたっては適切な条件であるが、その一方で、その変化率がさほど高くなかった場合、変化の内容を比較分析する際のサンプルが不足するおそれがある。

これを避けるため、評価対象文の予備候補として、前年度の二次評価にて作成した評価対象文を利用することとした。前年度の二次評価では、本評価の対象となる 174 テーマを含む多数のテーマコードについて、評価対象文を 10 文ずつ選定している。そこで、これらの評価対象文を「予備」文とし、こちらについても新規の機械翻訳結果を取得して、新旧で変化が生じているものは、1.5.3 にて選定した「新規」の評価対象文 522 文のうち新旧機械翻訳結果で変化がなかったものと差し替えて比較評価の対象とすることとした。こうすることで、新旧で機械翻訳内容が変化している文をより多く評価対象とすることができる。

「予備」文は、前年度の評価対象であるため基準翻訳文が作成されており、かつ文中の「重要技術用語」についても、基準翻訳文から抽出された適訳語が示されている。そして、当時の機械翻訳結果（以下「旧 MT」という。）でこれら重要技術用語が適訳語でなかった場合には、適訳語を前年度辞書データに加えている。その結果、「予備」文では、新旧の機械翻訳内容が（良い方向に）変化している確率が、無作為抽出である「新規」文に比べて高くな

るものと予測される。この点は、分析の際に十分考慮する必要がある。

1.5.5 新規機械翻訳文の取得

上記 1.5.3 にて新規に選定した評価対象文 522 文（174 テーマ×3 文）と、予備候補として前年度二次評価対象文からピックアップした 1,740 文（174 テーマ×10 文）について、特許庁にて新規の機械翻訳結果を取得した。

その結果、「新規」文 522 文では 343 文（65.7%）において、新旧の機械翻訳結果に有意な変化⁵が生じていた。一方、「予備」文 1,740 文では 1,536 文（88.3%）において何らかの変化が生じていた。

表 1-44：新旧機械翻訳における翻訳結果の変化割合

	総数	変化あり	変化なし
新規文	522 文	343 文 (65.7%)	179 文 (34.3%)
予備文	1,740 文	1,536 文 (88.3%)	204 文 (11.7%)

上表のとおり、「予備」文における変化率は「新規」文における変化率よりも有意に高いが、これは先述のとおり、「予備」文の場合、そこに含まれる重要技術用語が前年度辞書データに含められていることが多いためである。したがって、前年度辞書データの適用による翻訳結果の変化の度合いについては「新規」文での結果のみを参照するのが妥当であろう。なお、「新規」文のみに限定しても、変化の割合は 66.5%とかなり高く、全体の約 3 分の 2 で新旧の翻訳結果に変化が生じていることが判明した。

⁵ 新規文 522 文で新旧で変化のあった 344 件中、1 文（3G160）のみ、変化箇所が英数字記号の半角・全角の差異（「(4)」→「(4)」）のみであったため、実質的に変化なしとみなし、評価対象から除外した。

1.5.6 評価対象 500 文の選定

本評価の対象とする 500 文は、上記 1.5.5 にて新旧の機械翻訳結果に「変化あり」となった文を優先して選定しつつ、各テーマの文数が極力均等に（具体的には 2～3 文に）なるように選定した。具体的には、以下の優先順位とした。

- ① 新旧機械翻訳結果に有意な変化が生じていた「新規」文の全件（343 文）を採用した。これにより各テーマコードにて最大 3 文、最小 0 文が評価対象となった。

↓

- ② 上記①にて文数が 0～1 文にとどまった各テーマコードについて、当該テーマの「予備」文から新旧機械翻訳結果に有意な変化が生じた文を補充し、まずは全テーマとも極力 2 文となるようにした。

その結果、51 文が補充され、合計で 394 文となるとともに、全てのテーマにて 2 文以上を確保できた。

↓

- ③ 残りの 106 文については、現状で選定されている文が 2 文である各テーマについて、リストの上位から順に、(変化ありの)「予備文」から 1 件ずつを補充していった。

この結果、①～③で選定された 500 文の全てが、新旧機械翻訳結果に有意な変化が生じたものとなった。

下表に、評価対象 500 文における「新規」文と「予備」文の構成比率を示す。

表 1-45：新旧機械翻訳比較評価対象 500 文における新規／予備文の構成比率

新規文	予備文	合計
343 文 (68.6%)	157 文 (31.4%)	500 文 (100%)

また、テーマ単位での選定文数の状況は以下のとおりである。

表 1-46：新旧機械翻訳比較評価 評価対象文数別テーマ分布

評価対象文数	3 文	2 文
テーマ数	152 テーマ	22 テーマ

各テーマにおける評価文数および「新規」文と「予備」文それぞれの文数は、別添 1-10「新旧機械翻訳文比較評価対象テーマコード一覧」に示す。各テーマの評価対象文 500 文

の詳細については下記電子ファイルを参照されたい。

[『03 新旧機械翻訳文比較評価結果.xlsx』](#)

1.5.7 重要技術用語の選定

「重要技術用語の翻訳精度」の評価のための重要技術用語の選定についても、二次評価と同様の基準にて行った。詳細は 1.3.3 を参照されたい。なお、「予備」文については、前年度評価時に選定した重要技術用語をそのまま採用した。

本評価における重要技術用語 500 語の詳細は下記電子ファイルを参照されたい。

[『03 新旧機械翻訳文比較評価結果.xlsx』](#)

1.5.8 評価基準

本評価は、二次評価と同様に、各評価対象文に対して「内容の伝達レベル」と「重要技術用語の翻訳精度」の二つの観点から人手評価を行った。両観点ともに評価対象の分野に精通した二人の評価者に独立して評価させた点、2 名の評価者の平均をもって各文の評価スコアとした点も同様である。

「内容の伝達レベル」、「重要技術用語の翻訳精度」それぞれの評価基準等の詳細については、1.3.4 を参照されたい。

なお、本評価は、各文について、新旧の機械翻訳文を相対的に比較評価する点が二次評価と異なる。こうした相対的な比較評価の場合、比較対象となる新旧の機械翻訳文を同一の水準で評価することが非常に重要となる。このため、本評価においては、各文における新旧の機械翻訳文は、必ず同一の評価者が同時に評価するようにした。評価対象文の中に含まれる「予備」文は、前年度事業にて旧機械翻訳文は品質評価済みであるが、これらについても、前年度の評価結果は使用せず、本評価にて改めて新旧を同時に評価した。

1.5.9 新旧機械翻訳文の変化率の分析

1.5.5 に示したとおり、本評価では、「新規」文 522 文と「予備」文 1,740 文のそれぞれについて、新規の機械翻訳結果を取得した。しかしながら、その際に述べたように、前年度辞書データ約 10 万件の適用による新旧機械翻訳文の実質的な変化率の分析は「新規」文のみを対象とすべきであり、本項における分析もこの考えに基づき行う。

まずは「新規」文 522 文における変化率を下表に再掲する。

表 1-47：新規文 522 文における新旧機械翻訳結果の変化割合

	総数	変化あり	変化なし
新規文	522 文	343 文 (65.7%)	179 文 (34.3%)

上表のとおり、「新規」文 522 文における変化の割合は 65.7%と非常に高率であった。しかしながら、各文における機械翻訳結果の変化の全てが前年度辞書データの追加に起因するものとは考えにくい出力結果も散見される。

前年度辞書データの大半は専門技術用語（複数の語が連結されて名詞句となっている複合名詞。「环形风道／環状風路」など）であり、中国語 3 文字以上で構成される。以下に、前年度辞書データから用語が採用されたと見なせる機械翻訳結果の一例を示す。

(4E351-#2)

原文	本发明属于电路板领域，尤其涉及一种 薄膜线路板 及 薄膜线路板 的制造方法。
旧 MT	本発明はプリント基板分野に属し、特に 薄膜配線板 および 薄膜配線板 の製造方法に関する。
新 MT	本発明はプリント基板分野に属し、特に フィルム回路基板 および フィルム回路基板 の製造方法に関する。

上例は、前年度辞書データに含まれる複合名詞「薄膜线路板／フィルム回路基板」が適用されたことにより新旧 MT の機械翻訳結果に変化が生じたケースとみなせる。

なお、このように前年度辞書データから用語が採用された結果、文全体の構文解析結果が変化し、その他の（辞書データには含まれない）語の訳語も変化する場合もある。次に、これに該当すると思われる一例を示す。

(5H127-#1)

原文	氢能是一种公认的清洁 二次能源 ，其研究、开发和利用正受到越来越广泛的重视。
旧 MT	水素エネルギーは公認の 2回のエネルギー を洗浄することであり、それはちようどますます広範な重視を受けることを、開発研究し利用する。
新 MT	水素エネルギーは公認した清潔な 二次エネルギー であり、それはちようどますます広範な重視を受けることを、開発研究し利用する。

上例は、前年度辞書データから「**二次能源／二次エネルギー**」という語が採用された結果、この語を含む一節の構文解析結果が改善され、この語のみでなく、その周辺部分の訳文が「**2回のエネルギーを 洗浄すること**であり」から「**清潔な 二次エネルギー**であり」と変化している。

その一方で、前年度辞書データの適用が原因とは考えにくい変化も多数検出された。これを受け特許庁に確認したところ、新規の機械翻訳システムでは、前年度辞書データの適用以外にも各種の翻訳品質改善策が施されていることが判明した。特許庁の中日機械翻訳システムは、定期的に未知語対応・誤訳対応等のチューニングが施されており、今回の新規機械翻訳結果（以下「新 MT」という。）には、こうしたチューニングを経た最新のエンジンが使用されている。チューニングの内容に関しては非公開であるため断定はできないが、例えば下例などはこれに該当すると考えられる。

(2C030-#1)

原文	2. 准确规范：能够大大降低对笔画和字部安排的失准程度，提高汉字书写的规范性；
旧 MT	2. 正確に以下を規範する：ストロークと字部に 手配する 不整列程度を大幅に低下させることができ、漢字が書いた規範性を 高める ；
新 MT	2. 正確に以下を規範する：ストロークと字部に 配置する 不整列程度を大幅に低下させることができ、漢字が書いた規範性を 向上する ；

上例では、新旧の機械翻訳結果で変化した箇所は「手配する→配置する」、「高める→向上する」の2箇所であるが、先述のとおり、前年度辞書データの収録語は複合名詞のみであって、こうした一般動詞は含まれていない。

なお、先の例で示したように、文中の特定の語が前年度辞書データから採用された場合、（辞書データには含まれない）周辺の語にも変化をもたらす可能性はある。さらに、本例のように一見して複合名詞の訳語に変化が見られない場合でも、例えば前半の「失准程度／不

正列程度」という箇所が、旧 MT では「失准」と「程度」が別々に扱われていたのに対し、新 MT では「失准程度」という複合名詞として一語で取り扱われるようになった可能性もあり、こうした場合、この語自体の訳語に変化がなくとも、文全体の構文解析や訳語選定に何らかの影響を与えている可能性はある。

とはいえ、本例に限れば、この原文に含まれる 3 文字以上の複合名詞、具体的には「失准程度」と「規範性」のどちらも前年度辞書データには含まれておらず、これには該当しない。したがって、本例における「手配する→配置する」「高める→向上する」という訳語変化は、特許庁自身による定期的チューニングによるものとみなせる。

したがって、今回の新旧機械翻訳文比較評価は、前年度辞書データのみの適用効果を測るものではなく、前年度辞書データと定期的チューニングとを合わせた総合的な翻訳精度改善効果の測定という性質のものとなる。

1.5.10 新旧機械翻訳文の比較評価結果

ここまでに述べたとおりの方針・手順にて、評価対象文全 500 文（174 テーマ×3 文。ただし 22 テーマのみは×2 文）について、新旧機械翻訳文の比較評価を実施した。

本評価の評価基準及び評価スコアの集計方法は、すべて前出の二次評価と同様であり、各文について 2 名の評価者を起用して、5～1 の五段階評価による「内容の伝達レベル」と、A～D で評価する「重要技術用語の翻訳精度」の二観点からの評価を行った。なお後者については、集計にあたり A=5、B=3、C と D=0 とスコア換算した。

集計は、二つの観点それぞれについて 2 名の評価者の評価スコアの平均値を取得し、二つの平均値を取得して総合スコアとした。集計方法の詳細は 1.3.5 を参照のこと。

新旧 MT 比較評価の文単位の各観点の評価スコアの詳細については下記電子ファイルを参照されたい。

『03 新旧機械翻訳文比較評価結果.xlsx』

〔評価スコア平均値に基づく分析結果〕

まずは下表に、評価対象文全件（500 文）における「内容の伝達レベル」および「重要技術用語の翻訳精度」それぞれの平均評価スコアを示す⁶。

表 1-48：評価対象文全件 500 件における新旧 MT 比較評価結果

	旧 MT			新 MT		
	平均	(評価者 1)	(評価者 2)	平均	(評価者 1)	(評価者 2)
内容伝達レベル	2.76	(2.74)	(2.77)	3.18	(3.18)	(3.18)
重要技術用語	2.78	(2.73)	(2.83)	3.54	(3.50)	(3.58)
総合スコア	5.54	(5.47)	(5.60)	6.72	(6.68)	(6.76)

上表のとおり、「内容の伝達レベル」「重要技術用語の精度」とも、旧 MT よりも新 MT のほうが平均評価スコアが顕著に向上している。つまり、全体的に見て、新 MT は旧 MT よりも翻訳精度が改善していると見なせる。

ただし、今回評価対象とした文では、本評価のために新たに選定した「新規」文（343 文）と、前年度事業の評価に使用した「予備」文（157 文）とが混在しており、本評価における条件はそれぞれで異なる。このため、「新規」文と「予備」文とは、区別して分析する必要がある。

下表に、「新規」文と「予備」文、それぞれの評価観点別の平均評価スコアを示す。

表 1-49：新旧 MT 比較評価：新規文・予備文別集計結果

	内容伝達			重要用語			総合スコア		
	旧	新	UP	旧	新	UP	旧	新	UP
新規(343 文)	2.88	3.17	+0.29	3.06	3.34	+0.28	5.94	6.51	+0.57
予備(157 文)	2.50	3.21	+0.71	2.16	3.96	+1.80	4.66	7.18	+2.52
全件(500 文)	2.76	3.18	+0.42	2.78	3.54	+0.76	5.54	6.72	+1.18

さらに、「新規」文と「予備」文のそれぞれにおける、文単位での評価スコアの増減、す

⁶表中、「評価者 1」「評価者 2」欄には評価者別の評価値を示した。評価者 1、2 を担当した者は文によって異なるが、全体として 2 名の評価者間に大きな評価値の違いは見られず、おおむね一致した評価基準で評価が実施されたといえる。

なわち、新旧 MT で評価が向上した文数（表中「↑」）、下落した文数（「↓」）、変化しなかった文数（「→」）についても集計した。

表 1-50：新旧 MT 比較評価：文単位の評価スコア増減の状況

	内容伝達			重要用語			総合スコア		
	↑	↓	→	↑	↓	→	↑	↓	→
新規 (343 文)	126 (37%)	26 (8%)	191 (56%)	45 (13%)	18 (5%)	280 (81%)	136 (40%)	34 (10%)	173 (50%)
予備 (157 文)	105 (67%)	9 (6%)	43 (27%)	83 (53%)	8 (5%)	66 (42%)	116 (74%)	14 (9%)	27 (17%)
全件 (500 文)	231 (46%)	35 (7%)	234 (47%)	128 (26%)	26 (5%)	346 (69%)	252 (50%)	48 (10%)	200 (40%)

これら二つの表に顕れているとおり、「予備」文は「新規」文よりも新旧 MT 間での翻訳精度向上の度合いが顕著に高く、表 1-49 では「内容伝達レベル」では+0.71 ポイント（「新規」文は+0.29 ポイント）、「重要技術用語の精度」に至っては+1.80 ポイント（「新規」文は+0.28 ポイント）と飛躍的な向上を遂げている。また表 1-50 においても、「新規」文で総合スコアが「↑」、すなわち「新 MT>旧 MT」であった文は全体の約 40%であったのに対し、「予備」文では「↑」が約 74%という高率になっている。

ただし、前述のとおり、これには明確な理由がある。「予備」文は前年度事業にて翻訳精度評価に用いられた文を再利用したものである。そして前年度事業において、当時の評価実施時に各文から「重要技術用語」として選定され、かつ当時の機械翻訳（＝旧 MT）で適訳語に訳されていなかったものは、その大多数が前年度辞書データに適訳語（＝基準翻訳）で辞書登録されている。

したがって、この前年度辞書データを適用した新 MT において、「予備」文の重要技術用語が適訳語（すなわち A 評価）に改善し、結果的に「重要技術用語の精度」の評価スコアが大きく向上する確率が「新規」文に比べて顕著に高くなるのは当然のことといえる。事実、表 1-50 の「重要用語」の欄を見ると、「新規」文では「↑」（新 MT>旧 MT）が 13%にとどまっているのに比べ、「予備」では 53%と半数以上の重要技術用語において訳質が改善している。そして重要技術用語の精度向上は「内容伝達レベル」にもプラスの効果を与えるため、こちらの観点の評価スコアも連動して伸びることとなる。

具体例として、実際の「予備」文において、重要技術用語の訳語が前年度辞書データによ

って適訳語に変化し、結果的として新 MT の評価スコアが旧 MT に比べて大幅に向上したケースを示す。

(3C014-#3 評価者 1)

原文	进一步，所述底座上沿安装孔的周围设有<u>紧固套</u>，防止调节杆从安装孔中脱落。		
旧 MT	さらに、前記ベース上に取付穴に沿う周囲は<u>結合でカバーすること</u>を設け、調節レバーが取付穴中から脱落するのを防止する。		
	重要語：結合でカバーすること	用語スコア：C	内容スコア：3
新 MT	さらに、前記ベース上に取付孔に沿う周囲は<u>締付スリーブ</u>を設け、調節アームが取付孔中から脱落することを防止する。		
	重要語：締付スリーブ	用語スコア：A	内容スコア：4
基準訳	さらに、前記ベース上には取付穴の周囲に沿って<u>締付スリーブ</u>が設けられ、調節ロッドが取付穴から脱落するのを防止する。		

上例では、重要技術用語である「紧固套」が、旧 MT では「結合でカバーすること」という意味不明の誤訳となっていたところ、新 MT では前年度辞書データからの採用で「締付スリーブ」という基準翻訳どおりの適訳語に変化しており、結果として「重要技術用語の精度」の評価スコアが C (誤訳) から A (適訳) へ 2 段階改善している。また、これに伴い「内容伝達レベル」も 1 ポイント (スコア 3→4) 向上している⁷。

もちろん、「予備」文であっても、旧 MT の時点で重要技術用語が適訳 (A 評価) である場合もあるし、また様々な理由により重要技術用語が前年度辞書データに収録されていない場合もあり、全ての文で重要技術用語の評価スコアが向上するわけではない。ただし、前述のとおり、全体としては「予備」文は「新規」文に比べて圧倒的に有利な条件下にあり、前掲の表のとおり新 MT での改善度合いが高くなるのも当然といえる。

一方、「新規」文に関しては、選定にあたって文長をはじめとする種々の条件はあるものの、前年度辞書データの適用され易さ (例えば、前年度登録された用語を含む文を選ぶ等) という点に関しては一切の作為は行っていない。したがって、「予備」文よりも実態に近い条件、すなわち「中韓文献翻訳・検索システム」において現在実用されている機械翻訳から

⁷ 「締付スリーブ」の部分のみを見ればもう少しスコアが向上してもよい印象だが、別の箇所では「調節レバー」が「調節アーム」に変化しており、基準訳「調節ロッド」との合致度で新 MT の評価が若干下げられたものと思われる。

新エンジンによる機械翻訳に切り替えることで得られる改善の度合いがより正確にシミュレートされた評価結果であるといえる。

今回の評価では、「新規」文のみでの集計でも、新 MT の評価スコアは旧 MT に比べて有意に改善しており、特許庁による機械翻訳精度向上対策は確実な効果を挙げていると結論できる。

ただし、先述のとおり、「新規」文において（「予備」文においても、だが）、新旧 MT 間での翻訳結果の変化には、前年度辞書データの適用によるものと、特許庁が別途実施したエンジンのチューニングによるものがあり、一文において両者が同時に発生することも多い。したがって新旧 MT 間の評価スコアの差異には両者の影響が混在しており、スコア上で両者を区別して、個々に集計を行うことは難しい。

このため、次項からは、新旧 MT 間の変化の二つの要因である「前年度辞書データの適用」と「特許庁自身によるエンジンのチューニング」のそれぞれについて、それぞれに該当する実例を参照しながら、その改善効果と副作用について個々に分析することとする。まずは次項 1.5.11 にて、「特許庁自身によるチューニング」の効果について整理する。

（注記）

先出の二次評価では、技術分野（中 AU）ごとに平均スコアを取得して分野ごとの傾向を分析したが、新旧 MT 評価においては、一分野（テーマコード）あたりのサンプル数が 2～3 文と少ないこと、並びに、テーマにより「新規」文と「予備」文の構成比率が異なることから、有意義な比較にならないと判断し、今回の分析では割愛した。

1.5.11 比較評価結果の分析①「特許庁自身によるチューニング」

先述のとおり、本評価の対象とした新旧機械翻訳文において変化が生じた要因としては、
①前年度事業にて作成した用語辞書データによって技術用語の訳語が変化したケースと、
②特許庁自身による機械翻訳システムへのチューニングに起因するケースとがある。

後者②については、「特許庁により誤訳語・未知語等を収集し適訳語にて登録した辞書データ」に起因するものと、「エンジンのロジック自体のチューニングに起因する」によるものがあるが、いずれの場合も、前年度辞書の用語辞書データが中国語 3 文字以上の複合語のみを対象としているのに対し、より一般的な用語や表現が対象となっていることから、結果として、新旧機械翻訳文の変化の大半は、むしろこちら（特許庁によるチューニング）の要因によるものとなっている。

以下に、特許庁によるチューニングに起因すると思われる変化の実例をいくつか挙げる⁸。

(2D015-#2)

原文	所述 铲斗 或铲子前部边缘具有一个有斜面的唇缘或 刀刃 ，用于击打和穿透大量泥土和石块。
旧 MT	前記 铲斗 あるいはシャベル前方端縁は 1 個の斜面があるリップあるいは 刀刃 を有し、大量の泥と石塊を打ち突き抜ける。
新 MT	前記 バケット あるいはシャベル前方端縁は 1 個の斜面があるリップあるいは ナイフエッジ を有し、大量の泥と石塊を打ち突き抜ける。
基準訳	前記 バケット 或いはシャベルの前端には傾斜面のあるカッティングエッジ或いは ブレード を備え、大量の泥土或いは岩石を打撃する或いは突き刺すことに用いる。

上例 2D015-#2 では、原文の「**铲斗**」及び「**刀刃**」について、いずれも中国語 2 文字であり前年度辞書データには含まれていない。これらの語が、旧 MT ではどちらも中国語のまま出力されていたのに対し、新 MT ではそれぞれ「**バケット**」「**ナイフエッジ**」と訳出されるようになった。このうち前者（「**バケット**」）は人手翻訳による基準翻訳文（基準訳）と同じ適訳語である。後者（「**ナイフエッジ**」）については、この文ではやや不適訳であるが、少なくとも日本語に訳されているぶん、改善はしている⁹。

⁸ 特許庁による各種チューニングの詳細は明らかになっておらず、本分析の記載事項は提供された機械翻訳結果から推定したものである。

⁹ 中国語の場合、漢字であるため意味を類推できる場合があり、このため必ずしも言語のまま＝意味不明とはならないが、本来、機械翻訳で未知語を不適訳語よりも上位と評価するのは不適切である。

(5G352-#1)

原文	本発明涉及油气领域，特别是涉及一种 提高 高温埋地油气输送管道应力补偿器效能的方法。
旧 MT	本発明は油ガス分野に関して、特に埋めて油ガス搬送用導管応力補償器効能高温を 高める ことに関する方法。
新 MT	本発明は油ガス分野に関して、特に埋めて油ガス搬送用導管応力補償器効能高温を 向上する ことに関する方法。
基準訳	本発明は石油及びガスの分野に関し、特に地下の石油及びガス輸送パイプラインにおける応力補償装置の効能を 向上させる 方法に関する。

特許庁によるチューニングは技術用語に限らず、一般名詞や動詞についても行われているようであり、上例では「**提高**」という一般動詞の訳語が新旧 MT で変化している。なお、本件では旧 MT の「高める」と新 MT の「向上する」に優劣の差はない。

(3C061-#2)

原文	有鉴于此 ，本发明要解决的技术问题在于提供一种结构简单、切片整齐的手动切片器。
旧 MT	有してこれにかんがみて 、本発明が解決 した 技術的課題 が 構造が簡単であることを提供することにより、整然とした手動の切断器をスライスする。
新 MT	この点を踏まえて 、本発明が解決 する 技術的課題 は 構造が簡単であることを提供することにより、整然とした手動の切断器をスライスする。
基準訳	よって、本発明は技術的問題を解決するために構造が簡素で、スライス片が均等である手動スライサーを提供する。

先の 2 例が、特許庁によるチューニングのうち「誤訳語・未知語を登録した辞書データ」に起因するものと考えられるのに対し、上例は「エンジンのロジック自体のチューニング」によるものであろうと考えられる。具体的には「**有鉴于此**」という特許文献での定型的表現が妥当に扱われるようになったほか、「本発明が解決**する**技術的課題**は**」と、助詞の使われ方も適切なものに改善されている。今回の分析では、本例のように、特に特許文献での頻出表現へのエンジンによる対応が進んでいる印象を受けた。

なお、こうしたチューニング（辞書登録も含む）が文によっては裏目に出ることは機械翻訳では避けられない。当然ながら今回の分析においても、全ての変化箇所が改善と見なせたわけではなく、結果的に大差ないものや、むしろ悪化したものも少なからず検出された。こ

れらについて逐一の分析は避けるが、一点、悪影響の及ぶ範囲が大きく、次回以降のチューニングで見直すべきと思われる事象について報告する。それは、下例に示すように、ある種の文型において、本来あるべき助動詞「は」が欠落する事象である。

(3D203-#2)

原文	汽车顶盖是车身重要的覆盖件，其结构的好坏，直接影响车身的强度、刚度、NVH等方面的性能。
旧 MT	自動車ヘッドカバーは車体重要な被覆素子であり、その構造の良否は、車体の強度、剛性、NVHなどの方面の性能を直接影響する。
新 MT	自動車ヘッドカバーは車体重要な被覆素子であり、その構造の良否、車体の強度、剛性、NVHなどの方面の性能を直接影響する。
基準訳	自動車用ルーフは車体の重要なカバー部材であって、その構造の良し悪しは、車体の強度、剛性、NVHなどの性能に直接影響する。

上例では、旧 MT では「その構造の良否は」と妥当に訳されていたのに対し、新 MT では助詞「は」が出力されなくなり、結果としてこの部分の係り受けが不明瞭となってしまう。少なくともこの文に関しては明らかな悪化である。今回の調査では、この「は抜け」事象が、500 文中少なくとも 77 文にて検出された。

その中には下例のように、「は抜け」により日本語が自然になったものもあり、このようなケースを想定したチューニングであったのかも知れない。だが、今回の新 MT の該当 77 件¹⁰を見る限り、全体的には悪化とみなすべきケースが大半であり、チューニングの再考が必要と考える。

(2D026-#2)

原文	优选的是，所述的落叶处理装置中，所述收集管的另一端具有呈喇叭状的开口。
旧 MT	好ましくは、前記の落葉処理装置中は、前記収集管の他端はラッパ状開口部を呈することを有する。
新 MT	好ましくは、前記の落葉処理装置中、前記収集管の他端はラッパ状開口部を呈することを有する。
基準訳	好ましくは、前記落葉処理装置のうち、前記収集管の他端はラッパ状を呈する開口部を有する。

¹⁰ 該当文は電子ファイル『03 新旧機械翻訳文比較評価結果.xlsx』の欄外に「は抜け」=○として明示してある。

1.5.12 比較評価結果の分析②「前年度事業辞書データ適用による効果」

続いて本項では、前年度辞書データから訳語が採用されたことにより新旧 MT で種強く結果が変化したケースについて分析する。視認した限りでは¹¹、これに該当する文は 500 文中 186 文であった。

ただし、このうち「新規」文に該当するものは 89 文（343 文中。25.9%）であり、過半数は前年度評価に使用した「予備」文（157 文中 97 文。61.8%）が該当したものであった。

「予備」文の場合、適切に機械翻訳されなかった重要技術用語の多くが前年度辞書データに収録されているため、当然ながらここから訳語が採用される確率は高くなり、かつそれが適訳語であることも多くなる。以下、一例を示す。

（3B083-#2：予備文）

原文	図 7 是示出根据本发明的适用于 胶带清洁器 的引导槽与沿引导槽切开的胶带另一例的示例图。
旧 MT	図 7 は本発明による ゴムベルトクリーナー に適用するガイド溝ガイド溝に沿うことと切開するテープのもう一つの例の例示的図を示す。
新 MT	図 7 は本発明の 粘着テープクリーナー に適用するガイド溝 によって ガイド溝に沿うことと切開するテープのもう一つの例の例示的図を示す。
基準訳	図 7 は本発明を用いた 粘着テープクリーナー のガイド溝及びガイド溝に沿って切断された粘着テープの別の例を示した図である。

上例は、原文の「**胶带清洁器**」が旧 MT では「**ゴムベルトクリーナー**」という不正確な訳語で訳されていたところ、新 MT では前年度辞書データから訳語が採用され、「**粘着テープクリーナー**」という基準翻訳文どおりの適訳語に改善している。ただしこれは、前記したとおり、この評価文及び重要技術用語に基づいて「**胶带清洁器**／**粘着テープクリーナー**」という対訳語が前年度辞書データに登録されているため当然の結果といえる。

（なお、上例では新旧 MT で「によって」という部分にも変化が生じているが、これが前年度辞書データから複合語が採用されたことによる構文解析への影響なのか、特許庁によるエンジンのチューニングの結果なのかは判別できない。）

したがって、前年度辞書データの適用効果に関する分析は、「新規」文のみを対象とするほうが実態に近く適切である。以降の分析は、前年度辞書データから訳語が採用された 89

¹¹ 本分析は新旧機械翻訳結果の差分に基づいており、前年度作成辞書から採用されていたとしても新旧の翻訳結果が変わらない場合（例えば、これまで単語単位で翻訳されていた技術用語が複合語単位で訳されるようになったが、訳語としては変化がない場合。「統計」「機械」「翻訳」→「統計機械翻訳」など）は判別不可能である。

文について行う。

なお、「新規」文において前年度辞書データが適用された割合を、先ほどは 343 文中の 89 文＝25.9%と記載したが、これはあくまで「新旧 MT で変化のあった評価対象文」のうちで、その変化の要因が前年度辞書データであった文の割合であり、前年度辞書データの適用によって翻訳結果が変化する割合としては、「新規」の評価対象文全件すなわち 522 文を母数と見るべきである。つまり、522 文中の 89 文＝17.0%程度が、前年度辞書データから用語が採用されるカバー率と見なせる。

このカバー率については特段の基準値があるわけではなく、その多寡を評価することは難しいが、前年度辞書データは全て中国語 3 文字以上の複合語が対象であり、各語の出現率自体がそれほど高くはないと考えられること、及び、各文に対して約 10 万語の辞書データの全件が用いられるわけではなく、これを IPC 別に割り振った分野別辞書（詳細不明ながら、収録語数はかなり絞られると推測される）が、各文の筆頭 IPC に応じて個別に適用されていることを鑑みると、水準的な数値の範囲であろう。

次に、「新規」文にて前年度辞書データが採用された 89 文について、事例に基づきその改善効果を検証する。

ただし、この 89 文について、単純に新旧の評価スコアを比較することはできない。なぜなら、これら 89 文においても、前年度辞書データの適用に起因する変化だけでなく、特許庁自身によるチューニングに起因する変化が同時に発生していることが大半であるためである。

各文の品質評価は、文全体の内容伝達の度合いを評価した「内容の伝達レベル」と、文中から選定された特定の技術用語の訳質のみを評価した「重要技術用語の翻訳精度」の 2 つの基準にて実施しているが、前者においては当然ながら文中の全ての変化が対象となり、ここから前年度辞書データの評価のみを抽出することは不可能である。後者についても、前年度辞書データが採用された語が必ずしも今回重要技術用語に選定された語であるとは限らないため、同様である。

このため、これら 89 文について、本分析のために別途「前年度辞書データの適用により訳語が変化した語」のみをピックアップし、別途評価することとした。

具体的には、以下に示す A～C の 3 段階で評価した。なお、この評価は分析時に追加で実施したものであり、評価者は分析者自身である。

A：改善した

B：同等

C：悪化した

また、89 文中の 8 文に関しては、文中に前年度辞書データから採用されている語が 2 種類存在しており、これらについては 2 語それぞれを別々に評価対象とした。したがって、評価対象となった用語は全 97 語となる¹²。

下表に、89 文、全 97 語の評価結果の分布を示す。なお、各文における評価対象用語と評価結果は『03 新旧機械翻訳文比較評価結果.xlsx』の欄外を参照されたい。

表 1-51：前年度辞書データ採用用語 97 語の評価結果一覧

評価 A	評価 B	評価 C	合計
55 語 (56.7%)	36 語 (37.1%)	6 語 (6.2%)	97 語 (100%)

上表のとおり、前年度辞書データが採用された場合、その約 6 割において、少なくともその用語の訳質は改善されたと見なすことができた。以下、改善の主なパターンについて実例とともに示す。

〔評価 A：改善の主なパターン〕

①意味不明な訳語が改善されたもの

旧 MT では意味不明であった訳語が、前年度辞書からの採用により適訳に変化したケースである。旧 MT では正確な内容把握が難しいような訳語が、新 MT では適切な訳語に改善されており、少なくともその箇所においては翻訳品質が格段に向上している。

(2F044-#3)

原文	信号传输线缆 5 将所测得重量反馈至 称重仪表 6，各测量点 称重仪表 6 总和便是变压器重量。
旧 MT	信号伝送ケーブル 5 は検出した重量を 計量器表 6 にフィードバックして、各測定点 計量器表 は 6 の総和 は 変圧器重量である。

¹² 文中に同じ用語が複数回出現する場合は文単位で 1 語としてカウントした。つまり、同じ語が同一の文で何回出現しても 1 語としたのに対し、複数の文で出現する場合は、それぞれの文で 1 語として個別にカウントした。

新 MT	信号伝送ケーブル 5 は検出した重量を 秤量計器 6 にフィードバックして、各測定点 秤量計器 6 の総和変圧器重量である。
基準訳	信号伝送ケーブル 5 は測定された重量を 計量メーター にフィードバックし、各測定点の 計量メーター 6 の総和が変圧器の重量となる。
解説	原文の「 称重仪表 」(2 回出現) が旧 MT では「 計量器表 」と誤訳されていたところ、新 MT では前年度辞書の適用により「 秤量計器 」と妥当な訳語に改善した。 なお旧 MT では後半に存在していた助詞「は」が新 MT では 2 箇所で消失しており、1 つめは改善、2 つめは悪化とみなせるが、これが前年度辞書適用の間接的影響なのか、先に報告した「は抜け」の結果なのかは判別がつかない。

前年度辞書データが複合語のみを対象としているため、このデータから訳語が採用された場合、上例のように、旧 MT では単語ごとに訳されて単純に連結されていた（その結果意味不明となっていた）ものが、新 MT では一つの用語として適切に訳されるようになる、というケースが大半である。ただし、こうした中でも、旧 MT にて複合語の一部が名詞以外の品詞（とりわけ動詞）と解釈されているようなケースは、その時点で意味不明の訳文となるばかりか、文の構造も崩れてしまうことが多いため、辞書データ適用による改善効果はさらに大きくなる。

(2D012-#2)

原文	当 支撐油缸 4 的活塞杆伸长与缩短时，铲斗油缸 5 的活塞杆可以作等长的伸长与缩短。
旧 MT	シリンダ 4 のピストンロッドを 支持すること が伸長し短縮する時、バケットシリンダ 5 のピストンロッドは同一長さであることを作って短縮と伸長することができる。
新 MT	支持シリンダ 4 のピストンロッドが伸長し短縮する時、バケットシリンダ 5 のピストンロッドは同一長さであることを作って短縮と伸長することができる。
基準訳	支持シリンダ 4 のピストンロッドが伸縮する際、バケットシリンダ 5 のピストンロッドも同じ長さに伸縮することができる。
解説	原文の「 支撐油缸 」を旧 MT では複合語と解釈できず、「 支撐 」を動詞（支持する）と解釈し、「 油缸 」と別々に訳した結果、翻訳文中では両者が分離してしまい、原文の意味の類推が困難な内容となってしまった。新 MT では前年度辞書データから「 支撐油缸 」という複合語が採用され、適切な翻訳文が出力されている。

(4E004-#1)

原文	本发明涉及冶金设备和方法技术领域，尤其涉及一种自适应板坯 对中装置 及其 对中方法 。
旧 MT	好本発明は冶金設備と方法技術分野に関し、特に適応スラブの 対した中装置 およびその 対した中方法 に関する。
新 MT	本発明は冶金設備と方法技術分野に関し、特に適応スラブ センタリング装置 およびその センタリング方法 に関する。
基準訳	本発明は冶金設備と方法の技術分野に関し、特に自己適応式のスラブ センタリング装置 及びその センタリング方法 に関する。
解説	「 对中装置 」、「 对中方法 」とも、旧 MT では複合語として取り扱われず、「 対 」が「対する」という一連の用語として取り扱われず意味不明であるが、新 MT では前年度辞書の効果により複合語として取り扱われ、適訳語になっている。

(5E344-#2)

原文	本发明涉及电路板加工领域，具体而言，涉及一种 软硬结合电路板 的制作方法 and 用该方法制作的 软硬结合电路板 。
旧 MT	本発明は回路基板加工分野に関し、具体的には、 硬さがプリント基板と結合する 製造方法とこの方法で作製する 硬さに関してプリント基板と結合する 。
新 MT	本発明は回路基板加工分野に関し、具体的には、 リジッドフレキシブル回路基板 の製造方法とこの方法で作製する リジッドフレキシブル回路基板 に関する。
	本発明は回路基板の加工分野に関し、具体的には、 リジッドフレキシブル回路基板 の製造方法及び該方法で製造した リジッドフレキシブル回路基板 に関する。
解説	旧 MT では「 软硬结合电路板 」(2 回出現)を一連の複合語として認識できず、この語を主語 (软硬)・述語 (結合)・目的語 (电路板) からなる文と解釈した結果、意味不明になっている。新 MT ではこの語が複合語として取り扱われるようになったため、正しい訳語で訳されるようになり、さらに文末では、周辺部の構文解析にも良い影響を与え、「関する」の係り方が正しくなっている。

②訳語が一般的な用語に変化したもの

旧 MT においても意味はおおむね正しく理解できるが、ただし一般的な訳語ではなく不自然というケースである。

こちら、複合語が単語単位で別々に訳されていることに起因するという点では①と同様であるが、個々の単語の品詞解釈自体がおおむね妥当である場合、そこまで意味不明な訳語とはならないことも多い。ただし、特に技術用語においては、「意味さえ伝わればどう訳しても同等」ということはなく、一般的に用いられる定番の訳語が用いられるべきであることは言うまでもない。したがって、このようなケースも①と区別せず A 評価とした。

(2E125-#2)

原文	所述的磁吸盘，主要成分为 强磁铁 ，外观呈 圆台状结构 ，圆台中间有内螺纹，内螺纹两侧各有一个预留小孔。
旧 MT	前記の磁気吸着盤は、主要成分は 高磁界マグネット とし、外観は 円台状構造 を呈し、円錐台中間に雌ねじがあり、雌ねじの両側は1個のリザーブピンホールがそれぞれある。
新 MT	前記の磁気吸着盤、主要成分は 強力磁石 とし、外観は 円錐台状構造 を呈し、円錐台中間に雌ねじがあり、雌ねじの両側は1個のリザーブピンホールがそれぞれある。
基準訳	記電磁チャックは、 強力な磁石 が主な構成要素であり、外観は 円錐台形状 を呈する構造であり、円錐台中央に雌ねじを有し、雌ねじの両側に小さい穴がそれぞれ設けられる。
解説	「 强磁铁 」、「 圆台状结构 」が、旧 MT の「高磁界マグネット」「円台状構造」という、意味は把握できるが一般的には使われない用語から、新 MT の「強力磁石」「円錐台状構造」という定番訳語に改善している。 ※なお新 MT では冒頭に「は抜け」が発生しているが、これは前年度辞書適用の間接的影響ではないと考える。

A 評価のうち、②に該当するケースには、B 評価とすべきか微妙な、いわば境界線上にあるようなものも少なくない。例えば、「電子書籍」と「電子の書籍」とで優劣をつけるべきか、といった判断である。本分析では、このような「内容理解上は差はないが、普通は複合名詞で表現する」と判断される場合は A 評価とした。以下、これに該当する一例を示す。

(3L020-#1)

原文	为了增加塑料件 25 的耐磨性, 可以使用 耐磨性能 较好的材料制作, 例如 POM。
旧 MT	プラスチック片 25 の摩耗抵抗を増加するため、 摩擦に強い性能 の好ましい材料作製を用いることができ、例えば P O M。
新 MT	プラスチック片 25 の摩耗抵抗を増加するため、 耐磨耗性能 の好ましい材料作製を用いることができ、例えば P O M。
基準訳	プラスチック部品 25 の耐摩耗性を向上するため、 耐摩耗性 に優れた材料, 例えば P O M を使うことができる。
解説	原文の「 耐磨性能 」を、旧 MT では「 摩擦に強い性能 」、新 MT では「 耐磨耗性能 」と訳している。意味的にはどちらも同じことであるが、複合名詞で表現されるのが普通であるため、評価 A とした。

なお、上例からもわかるように、複合語として扱うべき語を単語単位でバラバラに訳すと、意味的には同じであっても、冗長な表現になったり、前後の語とのつながりが不自然になることも少なくない。

[評価 B : 実質的に変化なし]

前年度辞書データから訳語が採用され、新旧 MT で訳語が変化したが、内容理解のうえでも訳語の一般性においてもほぼ同等と見なされたケースである。ごく単純な例では、「ねじ」が「ネジ」に変化しただけというようなものも含まれる。以下、その代表例を挙げておく。

(3C061-#1)

原文	另外, 芒果果实所含有的粗纤维、蛋白质、矿物质等也是其 主要营养成分 之一。
旧 MT	また、マンゴ果実はある粗繊維、蛋白質、ミネラル物質などを含有して同じくその 主な栄養分 の一つである。
新 MT	また、マンゴ果実はある粗繊維、蛋白質、ミネラル物質などを含有して同じくその 主要栄養成分 の一つである。
基準訳	また、マンゴーの果実に含まれる粗繊維、タンパク質、ミネラルなども、その 主な栄養成分 である。
解説	原文の「 主要营养成分 」を、旧 MT では形容詞＋名詞「 主な栄養分 」、新 MT では複合名詞「 主要栄養成分 」で訳している。意味は同等、訳語の一般性にも大きな差が無く、優劣なし(評価 B)と判断した。

(5G045-#1)

原文	可選的，所述绝缘管为中空柱状，所述第一永磁体和所述第二永磁体与所述绝缘管的接触面的形状相适配。
旧 MT	選ぶ、前記碍子管は中空支柱状とし、前記第 1 の永久磁石と前記第 2 の永久磁石と前記碍子管の接触面の形状は互いに適応する。
新 MT	選択可能であり、前記碍子管は中空柱状とし、前記第 1 の永久磁石と前記第 2 の永久磁石と前記碍子管の接触面の形状は互いに適応する。
基準訳	任意選択的に、前記絶縁管は中空の柱状であり、前記第一永久磁石及び第二永久磁石における前記絶縁管との接触面の形状は一致する。
解説	原文の「中空柱状」を、旧 MT では「中空支柱状」、新 MT では「中空柱状」としている。実質的には同義語といえ、複合名詞「主要栄養成分」で訳している。意味は同等、訳語の一般性にも大きな差が無く、優劣なし(評価 B)と判断される。

上記各例をはじめ、評価 B に該当するケースは、少なくともその文に限れば「辞書データに含める必要のなかった」語とも言える。しかしながら、これは「この文において複合名詞を構成する各語の品詞解釈に成功した」結果であり、この品詞解釈は文を構成する各語の関係性において都度決定されるものであるから、ある文において複合名詞を構成する各語が適切に品詞解釈されたとしても、同じ複合名詞が他の全ての文でも常に正しく扱われるという保証はない。その点、複合名詞が一語として辞書登録されていれば、その語はほぼ必ず複合名詞として取り扱われるため、この語（を構成する個々の単語の一部）が誤った品詞解釈をされるリスクは非常に少なくなる。

また、仮に「各語が常時適切に品詞解釈される」ような語が存在したとしても、それはあくまで特定の機械翻訳システムでのことであり、異なるロジックを用いる他の機械翻訳システムでも同様かは保証されない。また、同一の機械翻訳システムであっても、エンジンのバージョンアップ等により品詞解釈のロジックが調整されることも多く、その結果、これまで適切に品詞解釈されていた語がそうでなくなるという可能性も否定できない。この観点から、特定のシステムの特定時点での出力結果のみで、専門用語辞書の要否を判断しないほうが賢明である。

これら 2 点から、今回の分析では評価 B となった語にも、機械翻訳辞書データとして収集した意義は十分にある。

下例は、原文の「**维护費用**」が、旧 MT では「**メンテナンス費用**」、新 MT では「**メンテナンス経費**」と訳されており、本項の評価では優劣なし（評価 B）としたものであるが、新旧 MT を比較すると、この「**メンテナンス費用／経費**」周辺、「節約する」にあたる部分の構文解析が大幅に良化していることがわかる。翻訳結果のみからは、これがどのような要因によるものかを特定することは不可能であるが、特定の語が複合名詞と品詞解釈されることにより周辺の品詞解釈が改善する可能性は大いに考えられ、本例がこれに該当する可能性もある。

(2E163-#2)

原文	用钢管制造公路桥梁不但可节省钢材、简化施工，而且可大大减少涂保护层的面积，节约投资和 维护費用 。
旧 MT	鋼管製造道路橋梁を用いることは鋼材を 節減 することができ、施工を簡略化するばかりでなく、保護層を塗る面積を大いに減少することができ て、節およそ投資とメンテナンス費用 。
新 MT	鋼管製造道路橋梁を用いることは鋼材を 節約 することができ、施工を簡略化するばかりでなく、 かつ 保護層を塗る面積を大いに減少することができ、投資と メンテナンス経費を節約する 。
基準訳	鋼管を用いて道路の橋梁を製造することは鋼材を節約し、施工を簡略化するだけでなく、保護層を塗布する面積を大幅に縮小し、投資コスト及び メンテナンスコストを節約する ことができる。

〔評価 C：悪化した事例〕

前年度辞書データから訳語が採用された結果、以前よりも悪い訳語になってしまったケースも若干数ではあるが検出された。以下、その例を示す。

(3L102-#1)

原文	本冰箱门 开启装置 60 固设于冰箱门 51 的下端，通过螺钉将壳体与冰箱门固接。
旧 MT	ねじによってハウジングと冷蔵庫扉固定接続を、本冷蔵庫扉 開口デバイス 60 は冷蔵庫扉 51 の下端に固設する。
新 MT	ねじによってハウジングと冷蔵庫扉固定接続を、本冷蔵庫扉 起動装置 60 は冷蔵庫扉 51 の下端に固設する。
基準訳	本冷蔵庫扉の 開口装置 60 は冷蔵庫扉 51 の下端に固定接続され、ネジによりハウジングと冷蔵庫扉が固定接続される。
解説	原文の「 开启装置 」は、基準訳のとおり扉を「開く」装置である。旧 MT ではこれを「 開口デバイス 」と、やや一般的でない表現とはいえ、正しい理解が得られる訳語に訳しているが、新 MT ではこの語が「 起動装置 」という不適訳語に悪化してしまった。「 开启 」という語は「 起動 」や「 オン 」という意味もあるため辞書データが間違っているわけではないが、この文に関しては不適であり、評価は C となる。

上例は、複数の訳し方がある語において、不適当なほうの訳語が使われてしまったパターンである。このように複数の（意味の異なる）訳し方があるような技術用語については、原則、主要な訳語を全て辞書データに含め、その上で各分野での出現頻度等で最も適訳となる可能性が高い訳語を機械翻訳辞書に採用するのが一般的である。ただし、この手法で選定した訳語はあくまで「全体として適訳となる可能性が最も高い語」に過ぎず、このようにして選んだ訳語であっても特定の文ではそれが不適訳となることもある。

機械翻訳辞書を用いる以上、意味の異なる複数の訳し方がある語においては、一定量の誤訳が生じることは現状の機械翻訳の宿命といえる。ただし、上記のようなケースであっても、「門（扉）」に続く「**开启装置**」であれば「起動装置」ではなく「開口装置」の意味であるとはほぼ確定できるため、「門开启装置」の単位で別途辞書データに（例えば「開扉装置」と）登録することで、実質的な訳し分けに相当する対処もある程度は可能である。

(3C030-#2)

原文	进一步，如图 1 所示，装配装置还包括一第一 上料装置 5 和一第二 上料装置 6。
旧 MT	さらに、図 1 に示すように、アセンブリ装置はまた一第 1 の 材料供給装置 5 と一第 2 の 材料供給装置 6 を含む。
新 MT	さらに、図 1 に示すように、アセンブリ装置はまた一第 1 の 供給装置 5 と一第 2 の 供給装置 6 を含む。
基準訳	さらに、図 1 に示すように、組立装置はまた第一の 材料供給装置 5 及び第二の 材料供給装置 6 を含む。
解説	原文の「 上料装置 」の「 料 」には「材料」の意義があり、旧 MT の訳語「 材料供給装置 」のほうが、新 MT の「 供給装置 」よりも正確である。

上例は、原語に含まれているニュアンス、具体的には「材料」という部分が前年度辞書データの登録からは欠落してしまっているケースである。ただし、これはあくまで上記原文においての話であり、実際には「**上料装置**」はそのまま「供給装置」「フィード装置」等と訳されることも多い。また、たとえば「**上料装置**」の直前に供給される材料の具体的な名称が付加されているような場合（例えば「コンクリート材料供給装置」など）、むしろ「材料」という語は不適、冗長となることもある。

こうしたケースは、基本的には各訳語を出現頻度等で比較して優劣をつけるのが妥当と考えるが、本件のように、特定の用語が複数の訳語を有し、かつ一つの訳語が別の訳語の一部と同一である場合（つまり「供給装置」は「材料供給装置」の下 4 文字と同一）、単純な出現頻度カウントでは、確実に前者のほうが高頻度になってしまうので、これを考慮した優劣判定を行うことが望ましい。本事業の調達仕様では、年度内に作成する約 10 万語の辞書データ間においては、この点の調整（上のケースであれば「供給装置」の出現頻度から「材料供給装置」の出現頻度を除く）を行なっている。

最後に、残る 4 件の C 評価の実例を示しておく。

(2F044-#2)

原文	伴随着现代传感器技术、电子技术和计算机技术的 迅速发展 ，电子天平有了突飞猛进的进步。
旧 MT	現代のセンサ技術、電子技術とコンピュータ技術の 急速な発展 に伴って、電子天秤はすさまじい勢いで前進した進歩があった。
新 MT	現代のセンサ技術、電子技術とコンピュータ技術の 高速発展 に伴って、電子天秤はすさまじい勢いで前進した進歩があった。

基準訳	現代のセンサー技術、電子技術及びコンピュータ技術の 急速な発展 に伴い、電子天秤は目覚ましい進歩を遂げた。
解説	原文の「 迅速发展 」を新 MT では「 高速発展 」と訳しているが、旧 MT の「 急速な発展 」のほうが一般的な表現である。

(3L345-#2)

原文	本发明还提供一种冰箱，该冰箱包括 风道组件 ，该 风道组件 的结构可参照上述实施例，在此不再赘述。
旧 MT	本発明はまた冷蔵庫を提供し、この冷蔵庫は 風路アセンブリ を含み、この 風路アセンブリ の構造は上記のとおり、再びここでくどくど述べないことを参照することができる。
新 MT	本発明はまた冷蔵庫を提供し、この冷蔵庫は ダクトアセンブリ を含み、この ダクトアセンブリ の構造は上記のとおり、ここでもうくどくど述しないことを参照することができる。
基準訳	本発明はまた冷蔵庫を提供し、該冷蔵庫は 風路アセンブリ を含み、該 風路アセンブリ の構造は上述した実施例を参照することとし、ここでは説明しない。
解説	「 风道组件 」の「 风道 」には「 風路 」と「 ダクト 」のどちらの訳例もあるが、本件は冷蔵庫に関するので「 風路 」が妥当である。

(3D137-#1)

原文	视觉效果更好，使用时无论是白天还是黑夜，都能快速找到 救生圈 的位置，进而实现快速施救。
旧 MT	視覚効果はより良く、使用時いずれにせよ日中か夜中であり、全て 救命浮輪 の位置を素早く発見することができてさらに、素早く救助を実現する。
新 MT	視覚効果はより良く、使用時いずれにせよ日中か夜間であり、全て 浮き輪 の位置を素早く発見することができてさらに、素早く救助を実現する。
基準訳	視覚効果はより良好で、使用時は昼夜を問わず、いつでも迅速に 救命浮輪 の位置を発見でき、迅速な救助を行うことができる。
解説	原文の「 救生圈 」を旧 MT では「 救命浮輪 」、新 MT ではただの「 浮き輪 」としている。文脈を考慮すれば「 浮き輪 」だけでも「救命用」との理解は得られ、また何度も出てくる場合はシンプルで読み易くなると思うが、用語単位では旧 MT の「救命浮輪」のほうがベターと判断した。

(2D003-#1)

原文	本发明涉及 工程机械领域 ，尤其涉及一种单电机双轴输出的电驱动履带式推土机。
旧 MT	本發明は 産業機械分野 に関し、特に単一モーターの二シャフト出力の電気駆動ブルドーザに関する。
新 MT	本發明は 建設機械分野 に関し、特に単一モータの二シャフト出力の電気駆動ブルドーザに関する。
基準訳	本發明は 作業機械分野 に関し、特に 2 軸出力の単独モータ電気駆動のクローラ式ブルドーザに関する。
解説	「 工程机械领域 」は基準訳の「 作業機械分野 」が適訳であり、新 MT の「 建設機械分野 」は限定しすぎ。旧 MT の「 産業機械分野 」のほうがベターと判断した。

1.5.12 まとめ

今回の新旧機械翻訳文の比較評価では、現行の「中韓文献翻訳・検索システム」で提供されている中国文献の機械翻訳結果（旧 MT）と、同システムの機械翻訳エンジンに前年度事業で作成した約 10 万語の機械翻訳辞書データを補充し、かつ特許庁自身によるエンジンのチューニングを施した新システムによる機械翻訳結果（新 MT）とを比較し、その翻訳精度の改善度合いを調査分析した。

その結果、新 MT は旧 MT に比べ、全体的に翻訳精度が改善されていることが確認できた。前年度辞書データが適用された文はもちろんのこと、適用されていない文においても、エンジンのチューニング効果による一般用語の訳質改善や特許文の頻出表現への対応の改善などにより、約 40%の文にて翻訳精度の改善が見られた（表 1-47 参照）。

ただし、その一方で、前年度辞書データが適用された文は全体の 17%程度にとどまっており（1.5.12 参照）、前年度辞書データの規模ではまだ必要な技術用語を十分にカバーできていないことも判明した。これについては本年度事業にて同等規模の辞書データが補充される見込みであるが、今後も技術の進歩とともに新たな技術用語が発生することを考えると、継続的な技術用語対訳の収集は不可欠であろう。

なお、特許庁によるエンジンのチューニングも大きな効果を挙げていることが今回の分析で確認された（1.5.11 等参照）。こちらは主に一般用語や構文解析といった、分野を問わず多数の文に関係するものであり、専門技術用語の収集とあわせ、今後も継続的・定期的な実施が望まれる。

2. 特定テーマコードの中国特許公報の和文抄録の作成

機械翻訳文の翻訳品質評価に基づき、174 のテーマコードに属する中国特許文献 8 万件を選定し、和文抄録を作成した。以下、その概要を記す。

2.1 和文抄録の作成対象

和文抄録の作成対象テーマおよび作成対象文献の選定基準については、「1. 中韓文献翻訳・検索システムにおける機械翻訳文の翻訳品質評価」の「1.4 和文抄録作成対象案件の選定」に記載したとおりである。

この基準に基づき和文抄録作成対象とした全中 AU（49 分類）を別添 1-9「中 AU／テーマコード別 和文抄録作成件数一覧」に示す。なお、選定された 8 万件の詳細は電子ファイル『02 和文抄録作成対象 8 万件.xlsx』を参照のこと。

2.2 和文抄録の作成結果

和文抄録の作成結果を以下に示す。

作成件数	80,000 件
平均文字数（日）	2,045 文字
（請求項）	672 文字
（発明の目的又は効果）	195 文字
（実施例）	1,177 文字
最大文字数（日） CN102078696B	29,286 文字
最小文字数（日） CN101426155B	302 文字

2.3 特殊案件

和文抄録作成にあたり、中国特許文献に実施例が存在しない等、和文抄録作成にあたり定常の翻訳箇所選定が行えない案件については、その都度対応案を策定のうえ、特許庁担当者の承認を得て所定の対応を取った。

本年度は、別添 2-1「特殊案件対応一覧」に掲載した 24 文献について、作業員より報告を受け、別添 2-1 に記載の対応を取った。

3. 国際調査報告で引用された文献の全文翻訳文の作成

特許庁より貸与された「全文翻訳文作成対象文献リスト」に掲載された中国文献（国際調査報告で引用された中国文献のうち日本語及び英語で公開されたファミリー文献を有しない中国文献）、全 20,032 件について、全文翻訳文を作成した。以下、その概要を記す。

3.1 全文翻訳文の作成結果

全文翻訳文の作成結果を以下に示す。

作成件数	20,032 件
平均文字数（日）	12,570 文字
最大文字数（日） CN103459399A	146,206 文字
最小文字数（日） CN102057294A	248 文字

3.2 原文が 4 万文字以上の文献

全文翻訳文の対象となる中国特許文献の文字数が 4 万文字を超えている場合、原文を 4 万文字程度に短縮したうえで全文翻訳文を作成した。これに該当する案件は 111 件存在した。一覧を電子ファイル『[04 4 万文字超過文献一覧.xlsx](#)』に示す。

4. 対訳コーパスの作成

本事業にて翻訳対象とした中国特許原文と、作成した翻訳文（和文抄録、全文翻訳文）とを用いて中日対訳コーパスデータを作成した。以下、その詳細を示す。

4.1 対訳コーパスの単位

対訳コーパスは、中国特許原文データを以下の基準で分割した文単位にて作成した。

- ・文中の句点「。」で分割
- ・文中の改行で分割

なお、上記基準により、出願人が一文を意図的に改行している場合は、その改行ごとに1レコードとなる。出願人の意図的な改行は、請求項における要素列挙や発明の効果の列挙にて多用されるものであるが、改行の単位で事実上一文として扱うべき場合が多いこと、さらには、意図的改行を連結して一文とすると文長が長大となり、対訳コーパスとしての有用性が低下することなどから、上記基準を採用した。以下、出願人による意図的な改行の一例を示す。

本発明的优点在于：	←改行
1、制备的阳离子淀粉改性AKD为固体，便于储存和运输；	←改行
2、产品熔融后通过高速剪切乳化，……其施胶效果高于传统的AKD乳液。	←句点



連結せず、意図的改行の単位で対訳コーパス化

本発明的优点在于：
1、制备的阳离子淀粉改性AKD为固体，便于储存和运输；
2、产品熔融后通过高速剪切乳化，……其施胶效果高于传统的AKD乳液。

なお、出願人の意図的な改行ではなく、原文データ上の不備等により一文が複数の改行によって分割されているものも存在する（例えば、公報レイアウトの改行位置にて逐一改行マークが挿入されている）。こうした場合は一文に連結した。

步骤 7:支持虚拟 SIM 卡的新移动终端开机后, 通过短消息连接支持虚拟 SIM 卡的呼	←改行
叫服务中心的服务号码, 支持虚拟 SIM 卡的呼叫中心向移动终端发送短消息提	←改行
示, 请用户通过新移动终端输入获取原移动终端的账户信息的功能代码。	←句点



連結して本来の文単位で対訳コーパス化

步骤 7:支持虚拟 SIM 卡的新移动终端开机后, 通过短消息连接支持虚拟 SIM 卡的呼叫中心的服务号码, 支持虚拟 SIM 卡的呼叫中心向移动终端发送短消息提示, 请用户通过新移动终端输入获取原移动终端的账户信息的功能代码。
--

なお、こうした原文の分割・連結は、翻訳文の作成前に行い、全ての翻訳作業はこの単位で実施された。翻訳・校閲にあたっては、専用のインタフェースを用い、原文と翻訳文との対訳関係のズレを物理的に防止した。以下、その一例を示す。

所述方法包括以下步骤：	前記方法は以下のステップを含む：伝送する必要があるデータ量又はサービスタイプに基づいてアップリンク及びダウンリンクの帯域幅を決める。
根据需要传输的数据量或业务类型决定上下行的带宽。	

文区切りのズレによる空欄のアラートや、原文と翻訳文の句点の個数チェック等、原文と翻訳文の正確な文対応を作業インタフェース上で確保。

こうした措置により、翻訳が完了した時点で原文と翻訳文とが一文対一文の単位で完璧に対応づけられた対訳コーパスデータが自動的に完成することとなる。したがって本事業の対訳コーパス作成においては原文と翻訳文との対応付けのためのデータ解析処理（文アライメントツールによる）は一切行っておらず、原則として本事業で翻訳対象とした全ての文を余すことなく対訳コーパス化している。

4.2 対訳コーパスの作成対象

対訳コーパスは、原則として、本事業にて作成した全ての翻訳文（和文抄録、全文翻訳文）を対象とした¹³。ただし、以下に該当する文については、中日対訳コーパスという観点からは有用な対訳文とならないため、あらかじめ除外した。

[(i) 原文がイメージタグ、テーブルタグ、数式タグ、化学式タグを含むデータ]

原文中に化学式や表、数式などを示すイメージタグ、テーブルタグ、数式タグ、化学式タグが存在する場合は、翻訳文では、原則として該当箇所を【IMG】に置換している。（テキスト化可能な文字列はテキスト化する場合あり）。

こうした場合、【IMG】に置換するにせよテキスト化するにせよ、原文がタグデータである以上、翻訳文と原文とは正確な対訳データとはなりえない。このようなデータが対訳コーパスに含まれると、コーパスの使用時に悪影響を与えるおそれがある。このため、原文がイメージタグ、テーブルタグ、数式タグ、化学式タグを含むレコードは、あらかじめ除外した。

※上付・下付文字タグはあらかじめ翻訳文から除去する仕様につき除外対象とはならない。

[(ii) 原文、翻訳文が英数字のみで構成されるデータ]

数式や元素記号の列挙など、原文、翻訳文ともに中国語、日本語を含まない文は、中日の対訳コーパスとしては不要であるため、あらかじめデータから除外した。

[(iii) コーパスデータの項目区切り文字が含まれるデータ]

コーパスデータは各項目の内容を「|||」（バーティカルバー3つ）で区切る仕様のため、原文中に「|||」が存在する場合、区切り文字であるかの区別がつかないことから、あらかじめデータから除外した。

¹³ 全文翻訳対象文献に既存の「日本語要約」及び「発明の名称」が存在する場合、当該文献の「要約」及び「発明の名称」についてはコーパス作成の対象外とした。一文対一文での翻訳が必ずしもされている保証がないためコーパスの精度を低下させるリスクがあり、また要約及び発明の名称と同一／類似の文は明細書本文中にほぼ確実に再度出現するため、明細書本文中の記載から得られる対訳コーパスによりカバー可能であると判断した。

[(iv) 原文に不備があり翻訳文と対応していないデータ]

たとえば下例のように、原文の不備により翻訳文と正確に対応しないデータがまれに存在する。このようなデータについて、典型的なパターンによる検出や、翻訳者・校閲者からの指摘に基づく特定を行い、データから除外した。

以示例的方式， 错误！未找到引用源。 给出了本发性的真空紫外光谱检测装置示意图。	実施例により， 図 1 は本発明の真空紫外スペクトル測定装置の概略図を示す。
---	---

「**エラー！引用元が見つかりません**」という意味のアラート。翻訳文はPDFを参照して「**図 1**」と正しく訳されているが、対訳コーパスとしては原文と対応しておらず、除外すべきと判断。

4.3 対訳コーパスの作成件数

本事業にて作成した対訳コーパスデータの内訳を以下に示す。

和文抄録より	1,360,238 件
全文翻訳文より	2,661,929 件
合 計	4,022,167 件

なお、完成した対訳コーパスデータに対し、以下のダブルチェックを実施し、全てのレコードが一定の文字数比率の範囲内であることを確認した。

【対訳文の文字数比較によるチェック】

何らかの理由で文対応不良が発生した場合に備え、原文と翻訳文の文字数比較に基づくチェックを実施した。

具体的には、コーパスの各レコードについて、日本語と中国語の文字数の比率を算出し、両者が大きく乖離する文対や、双方の文字数が非常に多い文対の存在の有無と、存在した場合はその内容を確認した。

以下、文字数比率が最も甚だしかったレコードと、文字数が最も多かったレコードを示す。いずれも、内容的には正しいものとみなされた。

・[文字数比率が最も甚だしかったレコード（比率が小さいもの）]

最も文字数比率が小さいレコードは、日本語文字数と中国語文字数の比率 0.3（つまり、中国語文字数が日本語文字数の約 3 倍）であった。比率の小さいレコードについて抽出した一部の例を以下に示す。同様に比率の低いパターンをサンプル的にチェックしたが、いずれも下例同様、中国語原文と日本語文の対応に問題はなかった。

日/中比率	日本語文	中国語文
0.3	枕	一种枕头
0.3	化合物X..... 1 0 0	化合物 X.....100
0.4	図：	在附图中：
0.4	<注記>	<备注>

・[文字数比率が最も甚だしかったレコード（比率が大きいもの）]

最も文字数比率が大きいレコードは、日本語文字数と中国語文字数の比率 14.0（日本語文字数が中国語文字数の 14 倍）であった。このレコードの原文「即」は、前後に存在するイメージについての接続詞的に用いられているため、翻訳文では意味の通る意識がなされている。同様に、「以下のように～する：」「～は以下のとおりである：」のような文についても比率が大きくなる傾向が見られたが、類似するパターンにおいて特に問題のないことを確認した。その他、比較的比率の大きくなる傾向のあるレコードは、EUC 文字での置き換えが不可能な文字を含んでいる文が該当した。このような文は、翻訳、校閲時の基準によって当該文字を「*（Unicode U+文字コード）」で置き換えることとしているため比率が大きくなっている。

日/中比率	日本語文	中国語文
14.0	すなわち以下のとおりである：	即
13.0	一：十*（Unicode U+E816）（訳注：“有”の左上部分）*（Unicode U+4E06）（訳注：“石”の左上部分）厂*（Unicode U+8279）（訳注：草冠）*（Unicode U+E82C）（訳注：“昔”の上部分）*（Unicode U+4E30）（訳注：“三”と“ ”を重ねた文字）束痾痾；	一：十ナ厂卅卅丰束痾痾
11.0	以下のように設定する：	设

・[文字数が最も多かったレコード]

文献番号 CN101636187A の明細書の翻訳文において、日本語文字数が 9,352 文字、中国語原文文字数が 8,210 文字であった。この文献の当該明細書は一文として存在しており、物質名が列挙され、途中で意図的な改行も含まれていないため問題がないと判断した。その他、一定以上の文字数を含むレコードについてサンプル的にチェックを実施し、問題のないことを確認した。

[NM（日本語文数－中国語文数）が 1-1 でないもの]

本事業の仕様に従い、対訳コーパスの各レコードには、日本語文数と中国語文数とを N-M の形で表示している。対訳コーパスは原則として 1 文対 1 文で作成しているため、大半のレコードは 1-1 となるが、そうならないものが若干数存在するため、代表的なパターンを以下に示す。

なお、この文数カウントは、「1＋各文の文字列中に存在する句点の数」で機械的に計算した。但し、次の句点はカウントしない。

- ・文末尾の句点
- ・閉じ括弧¹⁴直前の句点
- ・原文最後の句点以降の文字が指定文字¹⁵のみ

・「n-1」パターン①

原文は 1 文であるが、長大のため日本語の読みやすさ優先や要素列挙の明確化のため、日本語が複数文で訳されている。

1. コリン性蕁麻疹を治療する漢方薬組成物であって、その特徴は、以下の成分及び重量部にて製造されることである：ケイガイ 5－12，ボウフウ 5－12，センタイ 3－10，オウバク 5－12，レンギョウ 5－15，生石膏 10－40，クジン 5－15，ハクセンピ 5－15，ショウマ 3－10，オウゴン 5－12，ショウジオウ 5－20，ボタンピ 5－15，カンゾウ 3－8。/// 上記各成分を漢方製剤の一般的な方法によ	1. 一种治疗胆碱能性荨麻疹病的中药组合 物，其特征在于，由下列组分和重量份配制而成：荆芥 5－12，防风 5－12，蝉衣 3－10，黄柏 5－12，连翘 5－15，生石膏 10－40，苦参 5－15，白藓皮 5－15，升麻 3－10，黄芩 5－12，生地 5－20，丹皮 5－15，甘草 3－8，上述各组分按中药制剂常规方法制成汤剂、冲剂、丸剂或口服胶囊任
--	--

¹⁴ 丸括弧、かぎ括弧、角括弧、黒墨括弧、ダブルクォーテーション

¹⁵ 句点、読点、カンマ、セミコロン、コロン、ピリオド

り湯剤、沖剤、丸剤又は経口カプセルのいずれかにする。	一種。
----------------------------	-----

1. 点滴静脈注射に起因する静脈炎を治療する塗り薬であって、重量百分率で以下を含む：チユ 12-16%、ジチャ 12-16%、ヒョウヘン 4-7%。/// その特徴は以下のとおりである：該塗り薬は重量百分率でさらに以下を含む：ダイオウ 12-16%、セッコウ 12-16%、モツヤク 8-12%、カンスイセキ 12-16%、オウバク 12-16%。/// 各原料薬の重量百分率の合計は 100% であり、合わせて粉碎し、微粉にして 80 メッシュのふるいにかけて均一に混合することで配合され、均一に混合した後にクリーンボックス内に入れて使用に備える。	1. 一种治疗静脉输液致静脉炎的外敷药，按重量百分比包括有：地榆 12-16%、儿茶 12-16%、冰片 4-7%，其特征在于：该外敷药按重量百分比还包括有：川军 12-16%、石膏 12-16%、没药 8-12%、寒水石 12-16%、黄柏 12-16%，各原料药的重量百分比之和为 100%，共研细面过 80 目筛混合均匀配制而成，混匀后置于清洁盒内备用。
--	---

・「n-1」パターン②

原文特有の「；以及」「在于：」「步骤：」「其中：」「如下：」等での意図的な改行により、日本語が複数文となっている。

第一回転アセンブリであって、該前アームに設けられ、且つ該手首を接続することにより、該手首を該前アームに対して第 1 回転軸方向に回転運動を行い、該第 1 回転アセンブリは固定ベース、第 1 モーター、第 1 減速機、伝達機構及び軸受け機構を含むもの。/// そのうち、	第一旋转组成，装设于该前臂体并连接该腕体，使该腕体相对于该前臂体以第一旋转轴向进行旋转运动，该第一旋转组成包括有固定座、第一马达、第一减速机、传动机构及轴承机构，其中，
---	---

1. フミン酸含有の緩効性、放出制御可能な肥料の生産方法であって、前記肥料は N、P2O5、K2O、有機質、フミン酸アンモニウム及びフミン酸ナトリウムを含み、各成分の含有量はそれぞれ以下のとおりである：N+P2O5+K2O ≥ 20%，有機質 ≥ 20%，フミン酸アンモニウム：15-20%，フミン酸ナトリウム：3-5%。/// その特徴は、該生産方法が以下のステップを含むことである：	1. 一种腐植酸缓、控释肥料的生产方法，所述肥料包括 N、P2O5、K2O、有机质、腐植酸铵和腐植酸钠，各成分含量分别为：N+P2O5+K2O ≥ 20%，有机质 ≥ 20%，腐植酸铵：15-20%，腐植酸钠：3-5%，其特征为，该生产方法包括以下步骤：
--	--

・「1-n」パターン①

原文は誤記により複数文となるが、本来は 1 文のため結合し、日本語は 1 文としている。

(1) 材料選択及び前処理：密度が 60-220 kg/m で、まっすぐで皮をむき易い植物の茎を選択し、皮付きなら皮をむき、葉付きなら葉を取り除き、小枝付きなら小枝を取り除き、真っ直ぐにカットして、根部をカットしてバーを残す；	(1) 选料与预处理：选密度为 60-220kg/m。/// 的、通直易去皮 的植物茎杆，未剥皮的要剥皮，有叶的去叶，有小枝的去小枝，截弯取直，截根留杆；
--	--

・「1-n」パターン②

原文は複数文であるが日本語の読みやすさを優先し、文を結合して、翻訳文は 1 文としている。

16、請求項 1-7 に記載の装置を使用し、泡で蒸気を調整し、液体を濃縮して結晶化させる方法であって、その特徴は以下のとおりである：加圧した熱気流が発泡液における泡生成器（9）を通過した後、大量の気泡（8）を生成し、気泡（8）は熱気流を包囲し、該気泡（8）が発泡液（7）を通過した後、大量の泡（16）を形成し、該泡（16）は熱気流を包囲して、熱気流は泡の内表面から泡を加熱し、泡（16）は“之”字型の泡通路（18）で、移動方向を継続的に変える移動過程において、泡の内部に旋回する熱気流を生成して泡の内表面の蒸発を加速し、泡の内部に過熱した飽和蒸気が迅速に充滿するようにする。	16、一种使用上述权利要求 1-7 中所描述的装置，用泡沫制取蒸汽，液体浓缩。/// 结晶的方法，其特征在于，加压的热气流通过发泡液中的产泡器（9）后，产生大量的气泡（8），气泡（8）包围热气流，该气泡（8）穿过发泡液（7）后，形成大量泡沫（16），该泡沫（16）将热气流包围，热气流从泡沫的内表面加热泡沫，泡沫（16）在“之”字型泡沫通道（18）内，不断改变运动方向的运行过程中，泡沫内产生旋转的热气流加速泡沫内表面的蒸发，使泡沫体内迅速充满过热的饱和蒸汽。
--	---

該シンナムアルデヒドのナノエマルジョンを冷蔵庫に入れて -4℃で一週間保存した後、室温に戻す。	将该肉桂醛纳米乳在冰箱中 -4℃保存一周后。/// 恢复至室温。
--	---

・「n-n」パターン

括弧¹⁶内の句点は文の区切りとはみなすべきでなく、原則、括弧内の句点での分割は行なわれないが、機械的カウントでは文末尾の句点と閉じ括弧直前の句点以外の句点は一律でカウント対象となるため、nが2以上とされるケースが発生する。

本実施例において、本発明の車搬送用フレーム10に、昇降機構5が更に設置され、該昇降機構5は、車輪ホルダ3の高さ位置を調整するために用いることができ、且つ電動ホイストを選択する（当然ながら、手動ホイスト、油圧機構、空気圧機構又は他の機械的機構も好ましい選択肢である。///且つ車搬送用フレームに昇降機を設置しなくてもよく、外部の昇降機構、例えばフォークリフト等を用いて自動車を支承した車輪ホルダの高さ位置を上昇させるか又は低下させてもよい）。	在本实施例中，本发明的运车框10，还设置有升降机构5，该升降机构5可以用于调节车轮支架3的位置高度，并且选用电动葫芦（当然，手动葫芦、液压机构、气动机构或其它机械机构也是很好的选择。///并且运车框还可以不设置升降机，而是采用外部的升降机构例如叉车等来提升或降低承载汽车后的车轮支架的位置高度即可）。
---	---

¹⁶ 丸括弧、かぎ括弧、角括弧、黒墨括弧、ダブルクォーテーション

5. 辞書データの作成

中国特許公報の和文抄録の作成及び全文翻訳文の作成で作成した翻訳文（和文抄録、全文翻訳文）と原文からなる対訳コーパスから対訳辞書データを作成した。以下、その概要を記す。

5.1 辞書データの作成手順

対訳辞書データの作成手順として、まず対訳コーパスを解析し、対訳辞書候補データを作成した。その後、特許庁より貸与された約 237 万語の対訳辞書データとの突き合わせを行い、重複しないデータのみを人手確認用辞書候補データとして作成した（179,572 件）。この人手確認用辞書候補データを頻度順に、採用語数（下表「対訳辞書として適切」に該当）が 10 万語に達するまで人手確認を行い、対訳辞書データを作成した。人手確認結果を以下に示す。

表 5-1：対訳辞書データ候補 人手確認結果

対訳辞書として適切	100,000 件
見出し語自体が不適切	12,504 件
訳語自体が不適切	3,795 件
見出し語と訳語とが対応していない	3,549 件
辞書に登録する用語として不適切	5,692 件

また、上記 10 万語に加え、本事業にて実施した翻訳品質評価の対象とした重要技術用語のべ 6,500 語（二次評価：6,000 語、新旧 MT 評価：500 語）のうち、機械翻訳結果が基準翻訳文の訳語と異なるものについても、上記データとの重複削除を行ったうえで同様に人手確認を行い、辞書データとして適切と判断された 2,639 語を対訳辞書データに追加した。なお、これらの語は対訳コーパスからの採用ではないため、出現頻度が 0 の場合がある。

さらに、前年度事業辞書データのうち、その適用により新旧 MT 比較評価で「悪化（C 評価）」と判定された事例（1.5.12 参照）から、辞書データ自体に問題があったもの 7 語について、それぞれ以下のとおり、より一般的な訳語の追加登録、もしくは辞書データからの除去による対応を実施した。

表 5-2：前年度事業辞書データによる悪化事例への対応

中国語	日本語	対応
开启装置	開扉装置	追加：現行「起動装置」より一般的な訳語も登録
上料装置	材料供給装置	追加：同上。現行訳「供給装置」
风道组件	風路アセンブリ	追加：同上。現行訳「ダクトアセンブリ」
工程机械领域	作業機械分野	追加：同上。現行訳「建設機械分野」
迅速发展	高速発展	削除：訳語が不適切につき
救生圈	浮き輪	削除：同上。
工程机械领域	建設機械分野	削除：同上。

5.2 辞書データの作成件数

5.2.1 辞書データ作成件数

上記した、対訳辞書として適切と判断されたデータ（100,000 件）、評価データから採用した重要技術用語（2,639 件）、新旧 MT 評価での悪化事例への対応（4 件追加、3 件削除）、計 102,640 件と、特許庁より貸与された対訳辞書データとを統合し、「見出し語（中国語）、訳語（日本語）、品詞（見出し語（中国語））、品詞（見出し語（日本語））」が付された形式とした上で、UTX 形式に変換し、対訳辞書納入データ（納入物④）を作成した（2,471,760 件）。また、人手確認で不採用となった候補データ全件（25,540 件）についても、別途「人手確認により対訳辞書から除外したデータ」として納入した。

①対訳辞書データ（新規作成）	102,640 件
②特許庁辞書データ（既存）	2,369,120 件
対訳辞書納入データ（①＋②）	2,471,760 件

5.1.2 データフォーマット

以下、「対訳辞書納入データ」及び「人手確認により対訳辞書から除外したデータ」のフォーマットを示す。

5.1.2.1 「対訳辞書納入データ」フォーマット

ファイル名：JPO-CJ-DICT-29fy.utx

フォーマット：冒頭の#で始まる行はヘッダ部（UTX 1.11 に準拠）であり、その後のデータ部の各行は一つの見出し語；訳語のセットとなっている。

- ・ヘッダ部1

#UTX 1.11; zh-CN/ja-JP; yyyy-mm-dd; JPO

上記 yyyy-mm-ddには作成日が設定される。

- ・ヘッダ部2

src <tab> tgt <tab> ...

ヘッダ部2には下記データ部に示す記号項目名がタブ区切りで設定される。

- ・データ部

データ部には下記に示す項目がタブ区切りで設定される。

項番	記号項目名	項 目
1	src	中国語・見出し語
2	tgt	日本語・訳語
3	src:pos	中国語・品詞
4	tgt:pos	日本語・訳語
5	comp-num	複合語の語数
6	freq1-all	頻度情報・全分野
7	freq1-chem	頻度情報・化学分野
8	freq1-elec	頻度情報・電気分野
9	freq1-mach	頻度情報・機械分野
10	freq1-phys	頻度情報・物理分野
11	freq2-all	補正された頻度情報・全分野
12	freq2-chem	補正された頻度情報・化学分野
13	freq2-elec	補正された頻度情報・電気分野

14	freq2-mach	補正された頻度情報・機械分野
15	freq2-phys	補正された頻度情報・物理分野

なお、上記項目の複合語の語数、頻度情報、補正された頻度情報は「平成27年度 中国特許文献の機械翻訳の品質評価及び辞書整備に関する調査¹⁷⁾」の項番4.3.3 「統合された対訳辞書作成」に記載の作成方法に準拠し作成した。

5.1.2.2 「人手確認により対訳辞書から除外したデータ」フォーマット

ファイル名：cj_dict_del_1_4.txt

データ項目

下記に示す項目がタブ区切りで設定される。

項番	項 目	備考
1	人手確認結果	記号D1, D2, D3, or D4が設定される。 D1(見出し語自体が不適切) D2(訳語自体が不適切) D3(見出し語と訳語とが対応していない) D4(辞書に登録する用語として不適切)
2	中国語	
3	日本語	

¹⁷⁾ https://www.jpo.go.jp/shiryoutouchin/chousa/pdf/kikai_honyaku/h27_01.pdf

6. テーマコード情報の作成

本事業にて全文翻訳文を作成した中国特許文献 20,032 件の全件について、特許庁により機械的に付与されたテーマコードの妥当性チェックを実施した。チェックの結果に基づき、特許庁より貸与された「テーマコード情報リスト」の内容を更新した。以下、作業結果について報告する。

6.1 「テーマコード情報リスト」更新内容

テーマコードの妥当性チェックの対象となった中国文献 20,032 件における、機械付与テーマコードの総数と、このうち妥当（正解）であったテーマコード件数、誤りであり削除したテーマコード件数、不足であり追加したテーマコード件数について以下に示す。

	機械付与 テーマ数	正解 テーマ数	削除 テーマ数	追加 テーマ数	修正後 テーマ数
合 計	47,160	33,046	14,114	9,017	42,063
平 均	2. ³⁵⁴	1. ⁶⁵⁰	0. ⁷⁰⁵	0. ⁴⁵⁰	2. ¹⁰⁰
		70%	30%	19%	

上表に示したとおり、対象案件 20,032 件における機械付与テーマコード数は平均 2.354 であったのに対し、人手確認後の付与テーマ数は平均 2.099 と、若干（平均 0.255）の減少が見られた。

機械付与テーマコードの正解率は約 70%であった。また、一案件に対して機械付与されたテーマが全ての正解テーマ（＝確認後付与テーマ）を含んでいた案件は 20,032 件中 13,503 件（約 67%）であり、このうち 7,370 件（約 37%）は、確認後付与テーマと完全に一致した（つまり削除も追加も不要であった）。

なお、これらの集計結果はいずれも前年度（平成 28 年度）事業とほぼ同一の傾向であった。以下、前年度の集計結果を参考に示す。

（参考）平成 28 年度「テーマコード情報リスト」更新結果

	機械付与テーマ	正解テーマ	削除テーマ	追加テーマ	修正後テーマ
合 計	46,476	32,276	14,200	8,480	40,756
平 均	2. ³⁷⁸	1. ⁶⁵²	0. ⁷²⁷	0. ⁴³⁴	2. ⁰⁸⁶
		69%	31%	18%	