中国特許の請求項及び中国特許審決の
機械翻訳に関する調査事業

調査報告書

令和 5 年 3 月 24 日

一般財団法人　日本特許情報機構

内容

報 告 書 概 要 1
調 査 報 告 35
1. 本事業の概要 37
1.1. 本事業の背景・目的 37
1.2. 調査分析の内容 38
1.3. 調査対象 10 万件の選定基準 38
1.4. 調査分析の手順 39
1.4.1. 機械翻訳傾向の統計的な調査・分析 40
1.4.2. 現状の機械和訳における特徴的な不備の調査・分析 40
2. 機械翻訳傾向の統計的な調査・分析結果 41
2.1. 校正状況の記録 41
2.2. 重大な不備の判定精度の改善 41
2.3. 全請求項の集計結果と分析 43
2.4. 独立請求項・従属請求項別の集計結果と分析 45
2.4.1. 原文の平均文長と重大不備の発生率との関係の考察 46
2.4.2. 請求項内の意図的改行数に関する考察 46
2.4.3. 重大な不備の類型別の発生頻度の分析 47
2.4.4. まとめ 48
2.5. 請求項の文長に基づく集計結果と分析 49
2.5.1. 全請求項の文長別統計 49
2.5.2. 独立請求項／従属請求項の文長別統計 51
2.5.3. 意図的改行・区切り記号の有無 55
2.6. IPC セクション別の集計結果と分析 60
2.6.1. 全請求項の集計結果 60
2.6.2. 独立請求項／従属請求項別の集計結果 65
2.7. 出願人の国コード別の集計結果と分析 69
2.8. 出願人の法人／個人別の集計結果と分析 72
2.9. 英数字記号文字列を含む請求項の傾向 73
2.9.1. 10 文字以上の連続する英数字記号文字列を含む請求項 73
2.9.2. 数式／化合物名称を含む請求項 75
2.9.2.1. 10 文字以上の数式を含む請求項 76
2.9.2.2. 英数字で表記された化合物名称を含む請求項 78
3. 機械和訳における特徴的な不備の調査・分析結果 81
3.1. ニューラル機械翻訳方式の特徴 81

3.2. 用語の誤訳.....83
3.2.1. 学習データ中に存在しない語の誤訳.....83
3.2.2. 学習データ中に適訳語が存在する語の誤訳.....88
3.2.3. 係り受けの誤訳.....92
3.2.4. 事例集.....96
3.3. 訳抜け.....101
3.3.1. 用語レベルの訳抜け.....101
3.3.2. 入力文字数超過による訳抜け.....103
3.3.3. セミコロン区切り文での訳抜け.....107
3.3.4. 事例集.....110
3.4. 湧き出し.....112
3.4.1. 一般的な湧き出し.....112
3.4.2. 不要な特許用語の湧き出し.....115
3.4.3. 事例集.....116
3.5. 数値の処理.....119
3.5.1. 文末への数値出力.....119
3.5.2. 漢数字の処理に係る不備.....121
3.5.3. 漢数字を含む化合物名称における誤訳.....123
3.5.4. 漢数字「一」の処理.....125
3.5.5. 事例集.....126
3.6. 英数字記号の処理.....131
3.6.1. 数式.....132
3.6.2. 化合物名.....133
3.6.3. 事例集.....136
3.7. 請求項の記載様式に係る不備.....139
3.7.1. 請求項番号が本文の一部として扱われる不備.....139
3.7.2. 発明の主題を文末に出力する傾向.....142
3.7.3. 発明構成要素の訳ゆれ.....146
3.7.4. 長大な請求項での翻訳精度の低下.....149
3.7.5. 事例集.....150
3.8. サンプル 160 件における特徴的な不備の発生状況一覧.....152
3.9. 他種のニューラル機械翻訳エンジンとの翻訳結果比較.....155
3.9.1. 比較に使用したエンジン.....155
3.9.2. 用語の誤訳.....158
3.9.3. 訳抜け.....159
3.9.4. 湧き出し.....160

3.9.5. 数値の処理 160
3.9.6. 英数字記号の処理 161
3.9.7. 請求項の記載様式に係る不備 161
3.9.8. まとめ 161
4. 本調査のまとめ 163
4.1. 機械翻訳プラットフォームの現状の翻訳品質について 163
4.2. 機械翻訳プラットフォームの特徴的な不備と対処策について 165
4.2.1. 学習データが最新の技術用語に対応できていない 165
4.2.2. 前後処理を含めた長文の処理精度が不十分である 166
4.2.3. 数字（漢数字を含む）の処理方式に改善の余地がある 167
4.2.4. 請求項冒頭の請求項番号が本文に組み込まれて誤解釈される 167
4.3. 現状では改善困難な不備 168
4.3.1. マイナーな技術用語や出願人による造語の誤訳 168
4.3.2. 同一の技術用語の訳ゆれ 168
4.4. IPC セクション別／出願人の国コード別の翻訳精度の違いについて 169
添 付 資 料 171

報 告 書 概 要

Ⅰ．本事業の概要

本報告書は、2022 年 4 月から 2023 年 3 月まで 1 年にわたり実施した「中国特許の請求項及び中国特許審決の機械翻訳に関する調査事業（以下「本事業」という。)」の調査報告書（概要版）である。本章で、本事業の概要についてまとめる。

Ⅰ.1. 本事業の背景・目的

中国において日本国企業が特許を効果的に取得し、特許紛争に対して適切な対処をするためには、特許の権利範囲及び審判事件の傾向を把握する必要がある。すなわち中国特許の請求項及び中国特許審決の照会環境整備が急務である。

特許庁は、機械翻訳プラットフォームにより中国特許公報全文の機械和訳を作成し、特許情報プラットフォーム（J-PlatPat）を通じて提供している[1]。また、中国特許審決の機械和訳に特化したニューラル機械翻訳エンジンの学習用対訳コーパスの作成事業を実施した[2]。

中国特許審決については、上述した事業で作成された学習用対訳コーパスを活用した、高品質な全文機械和訳を速やかに作成して提供することが望ましい。

他方、中国特許の権利範囲の把握に関しては、権利への抵触防止等の観点において、その内容を正確に理解する必要がある。しかし、上述した現在提供中の中国特許公報全文の機械和訳にも、一定の誤訳が混入している。したがって、権利範囲を定める請求項の機械和訳については、調査・分析をして誤訳傾向を明らかにすることで品質向上に取り組むことが望ましい。誤訳傾向の調査・分析は、機械和訳に誤訳が含まれているかどうかを解析し、その誤訳がどのような校正を行うことで解決する誤訳であるか、その傾向を調べることによって行うことができる。調査・分析の過程で作成される人手校正文は、高品質な和訳として提供することが望ましい。

本事業では、上記した学習用対訳コーパスを活用した中国特許審決の高品質な全文機械和訳を作成した。具体的には、特許庁の作成した中国特許審決由来の学習用対訳コーパスを一般財団法人 日本特許情報機構（以下、「当財団」という。）保有のニューラル機械翻訳エンジンに学習させ、そのエンジンを使用して、6 万件の中国特許審決の全文機械和訳を作成

---

[1] 2015 年 1 月から、中韓文献翻訳・検索システムにより提供開始。2020 年 5 月以降は機械翻訳プラットフォームにより提供。

[2] 中国審決情報の機械翻訳に関する解析及びデータ作成事業(平成 30 年度から令和 3 年度)において、200 万コーパスを作成した。「みんなの自動翻訳@Textra®」におけるニューラル機械翻訳エンジンへ学習させることで、高い学習効果が確認されている。

した。

中国特許の請求項の機械和訳については、機械翻訳プラットフォームで提供されている10 万件を対象に、人手による校正の要否及び必要となる校正の程度に基づいて誤訳の有無とその内容を解析し、誤訳傾向を調査・分析した。これにより、権利範囲を定める中国特許の請求項のより正確な機械和訳を提供する上で、現在の機械翻訳プラットフォームが抱える問題の傾向（誤訳、訳抜け、湧き出しが発生しやすい傾向）を明らかにした。また、上記解析に伴って作成される中国特許の請求項の機械和訳の人手校正文を、高品質な和訳として提供する。これらを通じて、国内企業の中国における円滑な事業実施の支援を目指す。

Ⅰ.2. 調査分析の内容

本事業では、中国特許公報 10 万件の「特許請求の範囲」全文について、特許庁が機械翻訳プラットフォームで提供している機械和訳[3]を人手で校正し、発明を正確に理解できる品質の翻訳文を作成した。その上で、各案件に対して実施した校正の内容を解析し、現状の機械和訳における誤訳の発生頻度や傾向・特徴を統計的に把握するとともに、現状の機械和訳の特徴的な誤訳・訳抜け・湧き出しについて、実例に基づきその発生原因や解決策を分析した。本報告書で、その調査分析結果を報告する。

Ⅰ.3. 調査分析の手順

調査分析は、以下の２つの手法により行った。

Ⅰ.3.1. 機械翻訳傾向の統計的な調査・分析

調査対象の中国特許公報 10 万件の全請求項について、校正作業前後の和訳の内容を機械的に比較し、変化の有無（すなわち校正の有無）を取得した。さらに、校正時に重大な誤訳・訳抜け・湧き出しが検出された請求項に対しては、校正者自身がチェックマークを付し、発明の内容理解への悪影響が大きい重大な不備と、その他の比較的軽度な不備とを区別して集計できるようにした。

これら校正内容の記録を独立／従属請求項別、原文の文長別、技術分野別などの観点から集計し、機械翻訳プラットフォームにおける誤訳の発生傾向を統計的に調査・分析した。統計的調査・分析の結果は、第Ⅱ章にまとめる。

[3] 全件、現行のニューラル機械翻訳エンジンが用いられた 2020 年 3 月以降発行の公報から選定した。

Ⅰ.3.2. 現状の機械和訳における特徴的な不備の調査・分析

校正作業を通じ、現状の機械和訳における特徴的な不備について、多数の機械和訳に共通して発生しているものや、内容理解への悪影響の大きいものを中心に実例を収集し、分類した。そのうえで、不備の各類型について発生条件や発生原因などについて考察した。特徴的な不備の調査・分析の結果は、第Ⅲ章にまとめる。

Ⅱ. 機械翻訳傾向の統計的な調査・分析結果

本事業で校正した中国特許公報 10 万件の「特許請求の範囲」の全請求項について、校正の状況を記録・集計した。本章に、その統計的調査・分析の結果を記す。

Ⅱ.1. 全請求項と独立／従属請求項別の集計結果と分析

調査対象の全請求項 949,882 項における平均文長や校正の頻度、重大な不備の発生率[4]を集計した結果と、これらを独立請求項と従属請求項に区別して集計した結果を下表に示す。

表Ⅱ.1-1. 全請求項及び独立／従属項別の校正と重大不備の発生率

	全請求項	独立項	従属項
請求項数と構成比	949,882 項 (100%)	142,163 項 (15.0%)	807,719 項 (85.0%)
一請求項あたりの平均文長	138.4 文字	420.4 文字	88.7 文字
一請求項あたりの平均改行数	2.1 回	6.0 回	1.4 回
校正作業が発生した請求項	86.4%	99.5%	84.3%
重大な不備を含む請求項	21.4%	67.1%	13.4%
・重大な誤訳を含む	16.3%	49.7%	10.5%
・重大な訳抜けを含む	8.1%	35.6%	3.2%
・重大な湧き出しを含む	4.2%	19.0%	1.6%
・再機械翻訳を使用	0.4%	1.7%	0.2%

上表のとおり、本事業の校正作業では、全請求項の 86.4%に何らかの校正が加えられた。つまり、機械翻訳プラットフォームの機械和訳では、大半の請求項に何らかの校正すべき箇所が存在していた。一方、内容の正確な理解を妨げる重大な不備を含むものは全体

[4] 重大な不備の基準は「原文の技術的内容の正確な理解を大きく阻害するレベルの不備か否か」とし、この基準で各校正者に判定を行わせた。ただし、事業初期は判定に個人差が激しく、対策を施して差異を解消させた経緯がある。このため、本章で示す「重大な不備の発生率」は、個人差が十分に解消した事業後期分の数値を採用することとした。詳しくは本編第 2.2 項を参照。

の 21%程度にとどまった。

ただし、独立請求項と従属請求項に区分した集計では、両者の集計結果に大きな差異が見られた。特に、重大な不備の発生率は独立請求項が 67.1%、従属請求項が 13.4%と極端な差が生じている。これは、両者の文長（文字数）の違いによるところが大きいと考えられる。

請求項は、原則として一文で記載される。独立請求項は、発明の全体像を一文で詳しく説明するため、長大かつ複雑な文構造となる傾向が強い。一方、従属請求項では、発明の全体像は「請求項 n に記載の〇〇〇であって,」といった形で独立請求項を引用して簡潔に記され、さらなる限定部分のみが具体的に記載されるため、比較的シンプルな構造の短文となる。このため、独立請求項は 420.4 文字、従属請求項は 88.7 文字と、両者の平均文長には約 4.7 倍と大きな差が生じている。

機械翻訳は、文単位で翻訳処理を行う[5]。このため、一文で記載される請求項は、原則、項全体が一括で翻訳処理される。機械翻訳の難易度は文長が伸びるにつれて上昇するため、独立請求項が従属請求項に比べて不備が多くなるのは順当といえる。

Ⅱ.2. 重大な不備の類型別の発生頻度の分析

本調査では「重大な誤訳」、「重大な訳抜け」、「重大な湧き出し」、「再機械翻訳を利用[6]」の類型別に記録を取得した。本項で、独立請求項と従属請求項の不備の類型別発生頻度を比較する。

まず、「重大な誤訳」の発生率は、独立請求項で 49.7%、従属請求項では 10.5%であった。両者の比率は約 4.7：1 であり、文長比（4.7 倍）と合致している。

一方、「重大な訳抜け」と「重大な湧き出し」は、前者が従属請求項 3.2%に対して独立請求項では 35.6%（11.1 倍）、後者が従属請求項 1.6%に対して独立請求項では 19.0%（11.9 倍）と、いずれも 10 倍以上、文長比の 2 倍強に相当する差が生じている。つまり、独立請求項は、重大な訳抜け・湧き出しの発生率が従属請求項に比べて顕著に高い。この結果から、これら二種の不備は、入力文が長くなると発生率が大幅に高まる性質を有

---

[5] 一般的に中日機械翻訳では、文の認識を、主に句点（。）と改行コードにより行うと考えられる。改行を文区切りと認識することで、見出し文など文末に句点を伴わない文も適切に区切ることができる。なお、文が長大で一括で処理できる限度を超過する場合には、文を複数に分割して個別に翻訳することがある。

[6] 本事業では、機械和訳が著しく低品質で校正困難な場合に、別の機械翻訳エンジンによる再機械翻訳結果を利用可能とした。

すると仮定される。

「再機械翻訳の利用」については、独立請求項で 1.7%、従属請求項で 0.2%とともに少数ながら、両者の比率の差は 8.5 倍と開いている。この集計結果は、翻訳精度が極端に低いものが独立請求項に偏って存在したことを示している。

Ⅱ.3. 請求項の文長に基づく集計結果と分析

第Ⅱ.1 項では、独立請求項と従属請求項における重大不備の発生率の差異は、主に文長の違いに起因すると考察した。だが、独立請求項と従属請求項は記載様式にも違いがあり、文長以外の要素も不備の発生に関与している可能性もある。

これを検証するため、独立請求項と従属請求項をそれぞれ 50 文字単位で文長別に集計し、その結果を比較した。以下、独立請求項と従属請求項それぞれの、4 種の重大不備ごとの文長別発生率をグラフで示す。

図Ⅱ.3-1. 独立請求項の重大不備の類型ごとの文長別発生率

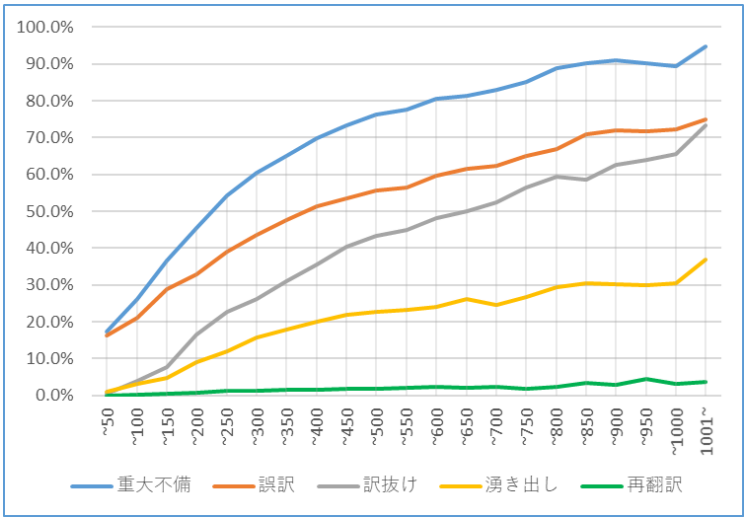
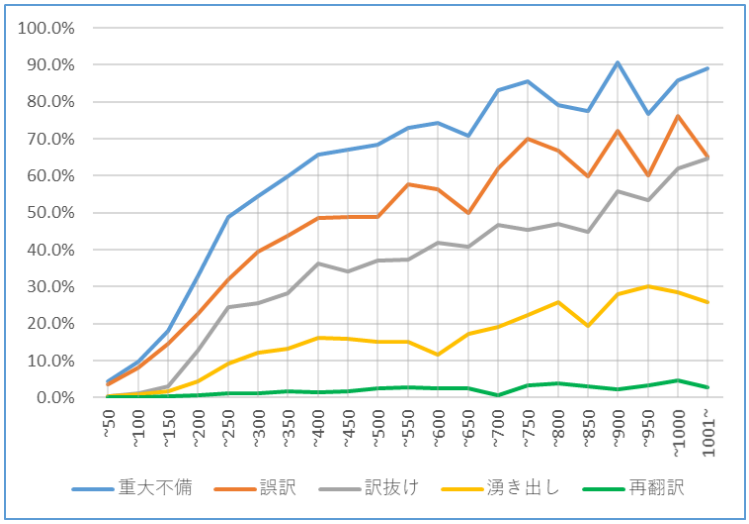


図 II.3-2. 従属請求項の重大不備の類型ごとの文長別発生率



上二図のとおり、独立請求項・従属請求項とも、重大不備の発生率は文長に相関して上昇している。ただし、重大な訳抜けと重大な湧き出しは、150 文字までは非常に低頻度で横ばいに近い推移であるのに対し、150 文字を超えた時点で不備の頻度が急増した。このことから、前項で述べた「これら二種の不備は、入力文が長くなると発生率が大幅に高まる性質を有する」という考察が裏付けられるとともに、「150 文字超」という具体的な文長閾値が示唆された。

上掲の二図では、独立請求項、従属請求項とも文長が伸びるにつれて重大な不備の発生率が増加しており、各文長範囲における発生率も大枠としては同程度といえる。したがって、第 II.1 項で考察したとおり、独立請求項と従属請求項の不備発生率の差は主に文長の構成比の違いに起因していることがわかる。

ただしその一方で、上二図では、文長 250 文字以下の区域では独立請求項のほうが明らかに不備の発生率が高く、その差は文長が短いほど顕著となることも見て取れる。これは、従属請求項の文型がパターン化されていることに起因すると考えられる。

従属請求項では、たとえば「根据权利要求 XX 所述的○○○，其特征在于，包括：…（請求項 XX に記載の○○○であって,さらに以下を含むことを特徴とする：…)」といった定番表現が案件単位で統一的に使用される。過去事例を学習して翻訳文を生成するニューラル機械翻訳方式は、定番表現を得意とする。このため、従属請求項は、確固とした定型パターンが存在しない独立請求項に比べて、同等の文長での翻訳難易度が低くなると考えられる。特に短文の請求項では、全体に占める定型パターン部分の割合が高まるため、不備の発生率が大幅に抑制されたと考えられる。

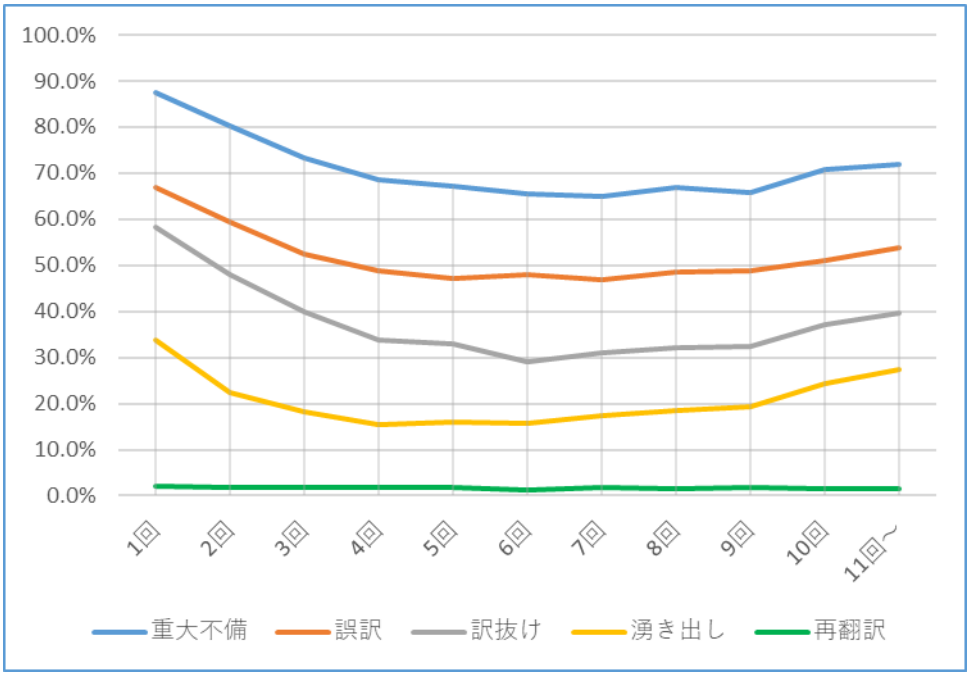
Ⅱ.4. 意図的改行・区切り記号の有無

機械翻訳処理の基本単位は文である。入力された文章は句点「。」で文単位に区切られ、個別に翻訳処理が行われる。また、文末の改行も区切りと見なされる。さらに、機械翻訳では、文が長大で分割処理が必要となるような場合、セミコロン「；」なども区切りとして扱われることが多い。

前項で示した文長別集計結果は、請求項全体の文長に基づく。請求項は原則として一文で記載され、末尾以外に句点が存在するケースは稀だが、長大な請求項には意図的改行やセミコロンが挿入されることが多い。よって、同じ文長の請求項であっても、意図的改行やセミコロンで適度に区切られた請求項と、こうした区切りが存在しない請求項とでは、実質的な翻訳難易度に差がある可能性が高い。

これを検証するため、独立請求項のボリュームゾーンである文長範囲 301～500 文字に属する 25,224 項をサンプルとして、請求項中の改行の回数別に不備の発生率を集計した。下図にその結果を示す。

図Ⅱ.4-1. 301～500 文字の独立請求項の改行回数別の重大不備発生率



上図のとおり、重大不備の発生率は改行数 1 回（＝請求項中の意図的改行はゼロ）の時が最多（87.6%）で、意図的改行数の増加とともに減少している。ただし、減少の度合いは改行数が増えるにつれて小さくなり、7 回で最少（64.9%）となった後は、むしろ微増傾向に転じる。この傾向は、不備の各類型でほぼ共通している。

本調査の結果から、文長 301～500 文字の範囲内の平均的なサイズの独立請求項においては、意図的改行が挿入されていない請求項に比べ、適度に意図的改行が挿入された請求項のほうが不備の発生が抑えられていると結論できる。この結果は予測されたとおりであるが、ただし、意図的改行によって翻訳処理単位の平均文長が数分の一となることに比べると、重大不備の減少度はさほど大きくはない。

調査ではさらに、301～500 文字かつ意図的改行ゼロの独立請求項 3,020 項を用いてセミコロンの有無、多寡の観点から同様の調査を行ったが、状況はほぼ同様であった。具体的には、セミコロンを全く含まない請求項の重大不備発生率は 96.4%であったのに対し、セミコロンが 6 個挿入された請求項では 72.3%まで低下したものの、その後は微増傾向に戻った。

これらの調査結果から、機械翻訳プラットフォームでは、長大な請求項を意図的改行やセミコロンで適宜区切って翻訳しており、それが翻訳精度に一定の効果をもたらしていることがうかがえる。ただし、独立請求項のボリュームゾーンにおける重大不備の発生率は、意図的改行やセミコロンが適度に挿入されているものでも依然 60%を超えている。これは、現行の機械翻訳プラットフォームにおいて、長文を改行やセミコロンで区切るという措置のみでは解決されない重大不備が多数存在することを意味する。第Ⅱ.2 項では、重大な訳抜け／湧き出しは入力文が長くなると発生率が大幅に高まる性質を有すると仮定したが、本項の結果から、その大きな理由として、これらの不備が意図的改行で適度な文長に区切られて翻訳処理されても発生率が大幅には下がらない性質を持つためであることが示された。

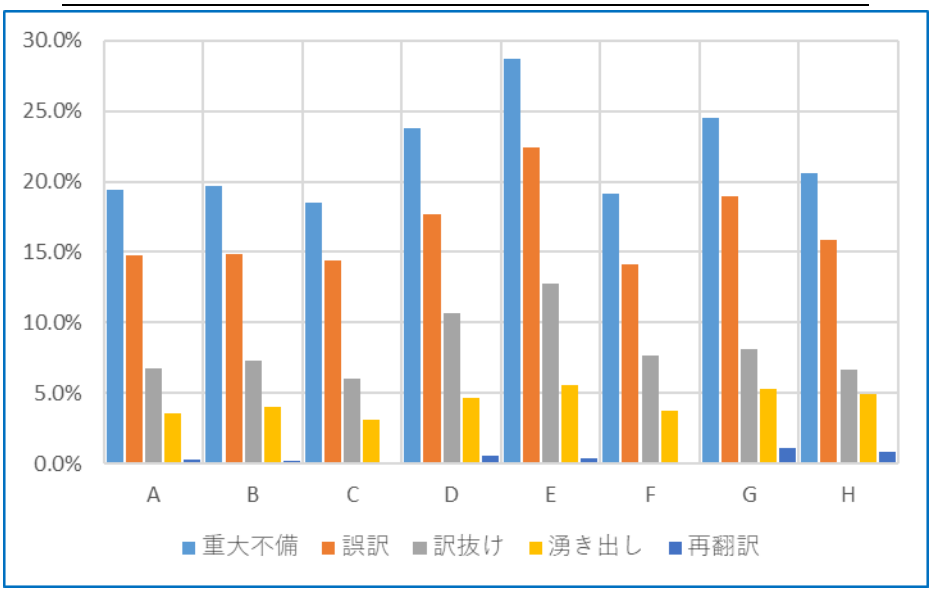
Ⅱ.5. IPC セクション別の集計結果と分析

中国特許公報には、発明の属する技術分野に応じて IPC（国際特許分類）が付与されている。本調査では、機械和訳における不備の発生状況に技術分野ごとの顕著な違いがあるかを把握すべく、全請求項を筆頭 IPC のセクション（A～H）別に集計・比較した。

比較を容易にするため、校正対象の公報 10 万件は、各セクションから 12,500 件ずつ、各セクションの対象案件の「特許請求の範囲」の平均文字数が 1,300 文字となるように選定している。ただし、そこに含まれる請求項の数や文長、独立／従属請求項の構成比等については特に調整していないため、各セクションの請求項数や平均文長には差異が生じた。こうした差異は、（平均文字数 1,300 文字という一定の条件下ではあるが）実際の中国特許の状況が反映されたものと見なせる。

以下、集計対象の全請求項（事業後期分 462,367 項）をセクション別に集計した結果を示す。

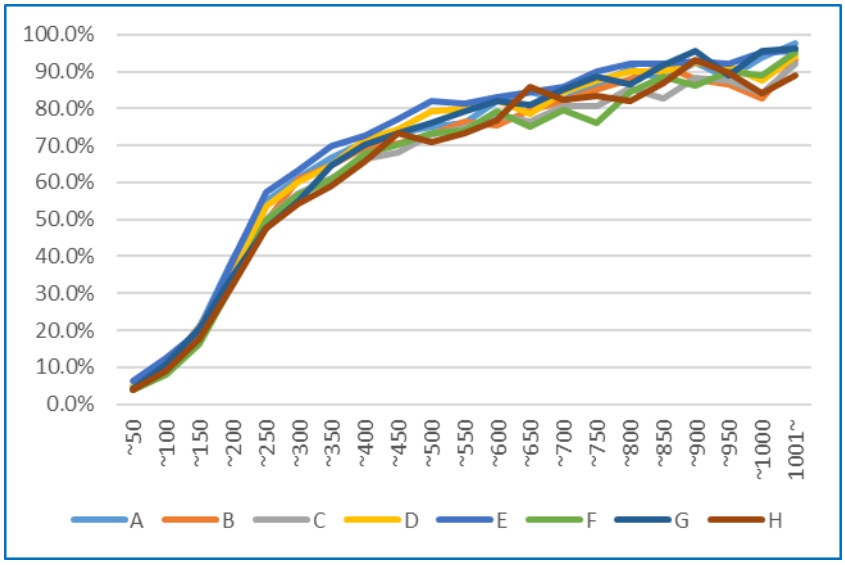
図 II .5-1. IPC セクション別 全請求項の重大不備の類型別発生率



各セクションの状況を比較すると、重大な不備の発生率は E セクションが 28.7%と最も高く、次いで G セクションが 24.5%、D セクションが 23.8%と続く。残る 5 セクションはいずれも平均値 21.4%を下回り、第 4 位が H セクションの 20.6%、第 8 位が C セクションの 18.5%と狭い範囲に密集した。不備の類型ごとの集計結果も、おおむね各類型とも E、G、D セクションの順に高く、他の 5 セクションは平均値を下回るという同様の傾向を示した。

E、G、D セクションで重大な不備が多発した最大の理由として、請求項の平均文長が他に比して長いことが挙げられる。具体的には、全セクションの平均文長 139.9 文字に対し、E セクションは 177.7 文字、G セクションは 151.7 文字、D セクションは 162.0 文字であった。一方、他の 5 セクションはいずれも全件平均文長を下回っていた。各セクションの重大不備発生率を文長別に比較しても、下図に示すとおり他から突出して高い／低い頻度のセクションは見られず、セクション間の不備発生率の差異が主に平均文長の差に起因することが裏付けられた。

図Ⅱ.5-2. IPC セクション×文長別 重大不備発生率



Ⅱ.6. その他の観点での集計結果と分析

本調査ではさらに、出願人の国コードに基づく中国出願／外国出願別、及び出願人の法人／個人別という観点から不備の発生頻度を比較したが、これらも IPC セクションと同様、文長差に準じた不備発生率となっており、文長範囲ごとの対比では差が見られなかった。

また、英数字記号文字列[7]や数式、化合物名称を含むものと含まないものとの比較では、同等の文長であれば、こうした英数字記号文字列を含むもののほうが不備の発生率が低かった。英数字記号文字列は原則として原文のまま機械和訳に転記すればよく、正しく取り扱う限り翻訳難易度は低い。本調査結果は、機械翻訳プラットフォームにおいて英数字記号文字列が適切に取り扱われていることを示すものである。

[7] 英数字記号は、請求項番号や句点、カンマ等を含めれば、全ての請求項に含まれている。ただし、長大な英数字記号文字列が含まれる請求項は、他の請求項に比べて翻訳精度に差異が生じている可能性が考えられる。このため、本調査では「10 文字以上の連続する英数字記号文字列」を調査対象とした。

Ⅲ. 機械和訳における特徴的な不備の調査・分析結果

本調査では、校正作業を通じて収集した現状の機械翻訳プラットフォーム機械和訳における特徴的な不備を、カテゴリ別に分析・考察した。本章にその要点をまとめる。

Ⅲ.1. ニューラル機械翻訳方式の特徴

機械翻訳プラットフォームは、ニューラル機械翻訳方式のエンジンを使用している。ニューラル機械翻訳方式は、エンジンにあらかじめ多量の学習データを学習させ、学習結果に基づいて最も確からしい翻訳文を生成する。

学習データには、原語と翻訳語による対訳文データ（対訳コーパス）が使用される。まず、この学習データをデフォルト状態のエンジンに学習させる。具体的には、対訳コーパスの各文をトークン（翻訳の最小単位）に分解し、トークン単位の対訳語データを生成するとともに、文中の他のトークンとの関係性を示すベクトル値を与える。そして、このベクトル値に基づいて学習データの各原文に対して機械翻訳を試行し、その結果を学習データの対訳文と比較して、学習データ全体の翻訳精度が最良となるようにベクトル値を調整していく。この処理を繰り返し、学習データ全体でのベクトル値が最適化された時点で、エンジンの学習が完了する。

学習が完了したエンジンに翻訳対象文が入力されると、エンジンは入力文をトークンに分解し、文中の周辺語の状況に基づいてベクトル値を付与する。そして、各トークンのベクトル値に近似する学習データ中の訳語候補をいくつか選定し、複数の翻訳文候補を生成して、最も確からしい文を採用する。このようにして翻訳処理が行われる。

この方式により、学習データ中に複数の訳語候補が存在しても、入力文の内容（すなわち周辺の語の状況）に応じて、その都度最適と判断される訳語が採用される。反面、学習データ中に存在しない用語は翻訳できず、訳語候補が存在しても、入力文のトークンと十分に近似なベクトル値でなければ（すなわち、周辺の語の状況がある程度類似していなければ）採用されない。

このように、ニューラル機械翻訳の翻訳精度は学習データの内容に大きく依存する。良好な翻訳精度を得るには、翻訳対象の文書と同種の対訳コーパスを大量に用意する必要がある。翻訳対象が特許文献である場合、学習データには主に対応特許公報由来の対訳コーパスが用いられる。

最後に、機械翻訳の処理単位について述べる。機械翻訳エンジンが１回の処理で取り扱

える文字数（トークン数）には上限がある。このため、機械翻訳では、入力された文字列を句点（。）や改行で文単位に区切る前処理を行い、一文ずつ翻訳処理を行っている。また、請求項で多用されるセミコロン（；）も、一文が長大な場合に適度な文長に分割する区切り位置として扱われることが多い。

機械翻訳プラットフォームのエンジンや前後処理、学習データ等の詳細は非公開だが、機械和訳の状況から、上記特徴の全てに該当すると考えられる。本調査は、これらニューラル機械翻訳方式の特徴を前提に、同プラットフォームの機械和訳に見られる各種不備を分析した。

Ⅲ.2. 用語の誤訳

ニューラル機械翻訳方式における用語の誤訳は、学習データ中に存在しない語が誤訳されるケースと、学習データ中に適訳語が存在するにもかかわらず誤訳されるケースとに大別された。

Ⅲ.2.1. 学習データ中に存在しない語の誤訳

ニューラル機械翻訳は、入力文を用語（トークン）単位に分解し、学習データ中に存在する語彙の中から、最も確からしい文を構成する訳語を採用する。このため、学習データ中に訳語が存在しなければ、正しく訳せない。該当する事例を示す。

［事例Ⅲ.2.1-1：CN107124391B ｾｸｼｮﾝ H］

原　　文	根据所述验证信息识别刷单行为，包括：识别所述无线接入点名称是否为随机字符串，若所述无线接入点名称为随机字符串，则确定所述客户端执行刷单行为；
機械和訳	前記検証情報に基づいてブラシ単体行動を識別し,以下を含む:前記無線アクセスポイント名がランダム文字列であるかどうかを識別し,前記無線アクセスポイント名がランダム文字列と呼ばれる場合,前記クライアントがブラシシング行動を実行するかどうかを決定する;
校正結果	前記認証情報に基づいて架空注文行動を識別することは,以下を含む:前記無線アクセスポイント名がランダム文字列であるか否かを識別し,前記無線アクセスポイント名がランダム文字列である場合,前記クライアントが架空注文行動を実行したと判断する;

本例原文の「刷单行为」は、ネット販売サイトで自社商品への架空の注文を（ロボットアクセスで）大量に行い、ランキングを向上させる行為を指す。近年使用され始めた新語である。訳語は校正結果の「架空注文行動」等が妥当だが、機械和訳では「ブラシ単体行

動」「ブラシシング行動」と大きく誤訳されている。

「刷单行为／刷单」という語は、過去の中日対応特許公報では使用例がなく[8]、機械翻訳プラットフォームの学習データ中にも存在しないと考えられる。学習データに存在しない用語は、出現の都度、必ず誤訳される。「刷单行为」は本案件（CN107124391B）の請求項中に計 8 回出現するが、訳語は上掲の「ブラシ単体行動」「ブラシシング行動」各 1 回、「ブラシ列」×4 回、「ブラシ動作」×1 回、そして訳抜け×1 回と、全て誤訳されている。

調査の結果、現行の機械翻訳プラットフォームでは「ディープラーニング（深度学习）」や「ビッグデータ（大数据）」、「充電スタンド（充电桩）」など、近年になり一般的に使われ出した新規の技術用語も誤訳されていることが判明した。具体的には、「ディープラーニング（深度学习）」は 184 項中 183 項（99%）、「ビッグデータ（大数据）」は 449 項中 408 項（91%）、「充電スタンド（充电桩）」は 141 項中 127 項（90%）と、ほぼ常時誤訳されていた。

こうした新語の誤訳は、学習データに訳例を追加することが唯一の解決策となる。新規の技術用語は、定着するにつれて特許公報での使用頻度も高くなる。したがって、最新の対応特許を定期的に学習させていけば、誤訳は高確率で解消する。

Ⅲ.2.2. 学習データ中に適訳語が存在する語の誤訳

ニューラル機械翻訳では、学習データ中に適訳語が存在しても、必ずその適訳語が採用されるとは限らない。主な原因として、用語が正しい単位で認識されないことと、訳語候補が複数存在し、入力文の分野や文脈に合わない訳語が採用されてしまうことが挙げられる。一例を示す。

［事例Ⅲ.2.2-1：CN113235177B ｾｸｼｮﾝ D］

原　　文	S3、通过侧吹风对复合纤维进行冷却成型，风温为 13-15℃，风速为 1-1.4m/s，风湿度为 80-90%,冷却吹风位置在喷丝板下方 40-60cm 处；
機械和訳	S3,サイドブローにより複合繊維を冷却成形し,風温は 13-15℃であり,風速は 1-1.4m/s であり,リウマチ度は 80-90%であり,冷却送風位置は紡糸口金の下方 40-60cm である;
校正結果	S3,サイドブローにより複合繊維を冷却成形し,風温は 13-15℃であり,風

[8] 当財団保有の、2021 年発行分までの中日対応特許のほぼ全件に由来する約 2.3 億文対の対訳コーパスを調査した結果である。以下同じ。

	速は 1-1.4m/s であり,風の湿度は 80-90%であり,冷却送風位置は紡糸口金から 40-60cm 下の箇所である;

原文の「风湿度」は、校正結果にあるとおり、ここでは「風の湿度」を意味する。したがて「风」＋「湿度」と分解するのが正しい。「风」、「湿度」とも一般用語であり、学習データ中に「風」、「湿度」と訳された事例が多数存在すること（つまり、本例の適訳語が存在すること）は確実である。だが、実際の機械和訳では「リウマチ度」と大きく誤訳されている。

この誤訳は、「风湿」が中国語で「リウマチ」を指すことに起因する。つまり、エンジンが入力文中の「风湿度」を「风」＋「湿度」ではなく「风湿」＋「度」と分解したことによる誤訳である。「风湿度」という複合語自体は中日対応特許公報に使用例がなく、機械翻訳プラットフォームの学習データでもこの単位ではトークン化されていないと推察される。このため、さらなる分解が行われ、学習データ全体のベクトル値から「风」＋「湿度」ではなく「风湿」＋「度」という解釈が優先されたと考えられる。

上例は用語レベルの局所的な分割誤りだが、より広範囲の誤解釈、例えば周辺の語のトークン分割の失敗や、文全体の分割の失敗によって、本来は正しく翻訳できるはずの用語が突発的に誤訳されることもあり得る。例えば、前掲の事例Ⅲ.2.1-1 で「刷单行为」に含まれる一般語の「行为（行動、行為)」がしばしば誤訳されたのも、このケースに該当する。ただし、このように用語が本来の単位で認識されず、意味の全く異なる訳語に誤訳されるケースは限定的であり、機械翻訳プラットフォームでは大半の語が、文の構造によらず[9]、正確にトークン分割されていた。

学習データ中に適訳語が存在するにもかかわらず誤訳が発生するもう一種の状況としては、有力な訳語が複数存在する用語において、入力文の技術分野や文脈に合わない訳語が採用されてしまうパターンが挙げられる。例えば「网」という語は、製紙分野ではパルプを載せる「網、格子、ワイヤ」を意味する語だが、通信分野では「インターネット」の略称として多用される。全技術分野を俯瞰すれば、「网」は「ワイヤ」や「網」よりも「インターネット」の意味で用いられることが多く、このため機械翻訳プラットフォームのように全技術分野を一括で扱うシステムでは、製紙分野の案件でも「网」が「インターネット」と訳されてしまうことが多くなる。

学習データ中に適訳語が存在する用語の誤訳は、特定の文におけるトークン分解の失敗

[9] ただし長大な文を除く。

や、ベクトル値の全体最適化の結果に起因している。このため、学習データにどのような訳例を追加すれば（あるいは除去すれば）誤訳が解消されるか、予測が困難である。また、例えば「网」を「網」や「ワイヤ」と訳した対訳文を集中的に学習させることで、製紙分野の案件でこの語が「インターネット」と誤訳される頻度を減らせる可能性はあるが、その結果、それまで「インターネット」と適切に訳されていた通信分野の案件で「網」や「ワイヤ」が採用されて翻訳品質が悪化する懸念が大きく、推奨できない。対処が難しいタイプの誤訳といえる。

Ⅲ.2.3. 係り受けの誤訳

係り受けとは、語同士の文法的な関係性を指す。誤訳は、用語の誤訳と、係り受けの誤訳とに大別される。ただし、ニューラル機械翻訳方式においては、どちらの誤訳も不適切なトークンが採用された結果であり、発生のメカニズムは共通している。本項で、係り受けの誤訳について分析する。まず一例を示す。

［事例Ⅲ.2.3-1：CN105905272B ｾｸｼｮﾝ B］

原　　文	上平坦部(31)和下平坦部(35)，所述上平坦部(31)和所述下平坦部(35)形成与所述底板(17)的所述后边界(19)相邻接的弯折部(41)，所述上平坦部(31)与所述中心纵向轴线垂直，所述下平坦部(35)与所述底板(17)形成包括在 30° 与 60° 之间的角度；
機械和訳	上平坦部(31)と下平坦部(35)であって,前記上平坦部(31)と前記下平坦部(35)は前記底板(17)の前記後境界(19)に隣接する屈曲部(41)を形成し,前記上平坦部(31)と前記中央長手方向軸に垂直であり,前記下平坦部(35)と前記底板(17)は 30° と 60° の間に含まれる角度を形成する;
校正結果	上平坦部(31)と下平坦部(35)であって,前記上平坦部(31)と前記下平坦部(35)は前記底板(17)の前記後縁部(19)に隣接する屈曲部(41)を形成し,前記上平坦部(31)と前記中心縦軸は垂直であり,前記下平坦部(35)と前記底板(17)は 30° と 60° の間に含まれる角度を形成する;

本例原文の「所述上平坦部(31)与所述中心纵向轴线垂直」は、校正結果のとおり「前記上平坦部(31)と前記中心縦軸は垂直であり」という意味であるが、機械和訳では「前記上平坦部(31)と前記中央長手方向軸に垂直であり」となっており、不正確である。文法的にも無理がある解釈で、典型的な係り受けの誤訳である。

ニューラル機械翻訳は、入力文をトークンに分解し、その単位で学習データ中から最適の訳語を選択する。ただし、トークンの単位は学習データ中で頻出の「文字列」であり、必ずしも文法上の「語」とは限らない。語よりもさらに細かく、文法的には区切れない単

位まで分割されることもあれば、複数の語が連なったフレーズ単位となることもある。かつ、中国語には「てにをは」に相当する助詞が存在しないため、中国語名詞に対して、日本語は「名詞＋助詞」の形でトークン化されることが多い。

前掲の事例においても、中国語「轴线」に対して、学習データ中には「軸に」「軸は」をはじめ「軸が」「軸を」などさまざまな訳語のトークンが存在すると考えられる。どれが採用されるかは、文法規則ではなく、ベクトル値に基づく。このため、入力文の内容や学習データの状況によっては、訳語の選択を誤り、結果的に非文法的な訳文が生成される場合もあり得る。上例はこれに該当するケースと見なせる。つまり、ニューラル機械翻訳における係り受けの誤訳は、学習データ中の訳語候補（トークン）から入力文にそぐわないものを採用した結果であり、本質的には用語の誤訳と同じ発生メカニズムといえる。

また、中国語を日本語に翻訳する際には、語順（トークンの順序）も並べ替える必要がある。ニューラル機械翻訳では、翻訳文の語順決定にも文法規則を用いず、学習したベクトル値により各トークンをさまざまな訳語と語順に組み替えて複数の翻訳文候補を生成して、最も確からしいと判断した一文を採用する。文法に基づいて翻訳文を生成していないため、文法規則に反する語順が採用される可能性もある。機械翻訳プラットフォームの機械和訳においても、不正な語順決定により係り受けの誤訳が生じている事例は多数検出された。

ニューラル機械翻訳における係り受け誤りは、用語の誤訳と同様、学習データの内容・状況に起因している。したがって、適切な学習データを追加することで改善される可能性がある。ただし、用語の誤訳に比べると、どのような学習データを追加すれば誤訳が改善されるかの予測が難しく、また、特殊な構造の文を過度に学習させると通常の文の翻訳品質に悪影響を及ぼす危険性もある。

Ⅲ.3. 訳抜け

訳抜けとは、原文の内容が翻訳文に反映されない不備を指す。主な原因として、原文中の用語や文言に対応する訳例が学習データ中に存在しないことや、入力文の不正なトークン分割により学習データ中の訳語が採用できなかったことなどが挙げられる。

訳抜けには、用語レベルのものから、意図的改行やセミコロンで区切られた翻訳処理単位レベルの大規模な欠落まで多岐にわたる。用語レベルの訳抜けは、学習データ中に存在しない難訳語において発生しやすい傾向はあるものの、一般語が突発的に訳抜けするケースも少なからず見られた。この状況は、前項に示した用語の誤訳の発生状況と共通する。事実、事例Ⅲ.2.1-1 の「刷单行为」をはじめ、多くの事例で同一の用語が文によって誤訳

されたり訳抜けしたりしていた。つまり、用語レベルの誤訳と訳抜けは、現象としては別種だが、どちらも翻訳不能な文言の存在によって生じる不備のバリエーションであり、根本的には同一の課題といえる。

一方、翻訳処理単位レベルの大規模な訳抜けは、第II章の統計的分析でも確認されたとおり、長大な請求項で多発する傾向が強かった。以下、一例を示す。

［事例III.3-1：CN106620893B ｾｸｼｮﾝ A］

原　　文	① 1.人工晶状体，其包括：
	② - 基体材料，选自疏水型丙烯酸酯、丙烯酸酯水凝胶、硅胶、硅水凝胶、氟硅丙烯酸酯、聚苯乙烯和聚甲基丙烯酸甲酯、聚碳酸酯、聚硅氧烷，或其混合物；
	③ - 至少一种光敏剂，选自光动力型和/或光热型光敏剂，光敏剂含有可聚合基团；
	④ 其中，光敏剂与基体材料以共价键方式结合，选自：
	⑤ - 光敏剂在基体材料成型过程中参与聚合；和/或
	⑥ - 光敏剂以表面接枝、表面修饰方式固定在基体材料表面，
	⑦其中基于材料总重量，所含光敏剂含量为小于 1 重量%。
機械和訳	① 1.眼内レンズであって,それは以下を含む:
	② -疎水性アクリレート,アクリレートヒドロゲル,シリカゲル,シリコーンヒドロゲル,フルオロシリコーンアクリレート,ポリスチレンおよびポリメチルメタクリレート,ポリカーボネート,ポリシロキサンからなる群から選択されるマトリックス材料;
	③ または光力学および/または光熱型光増感剤からなる群から選択される少なくとも 1 種の光増感剤;
	④ 光活性剤および/または光活性剤からなる群から選択される少なくとも 1 種の光増感剤;
	⑤ および/または
	⑥ -光増感剤が表面グラフト化され,表面修飾されて基材表面に固定される。
	⑦ この場合,含有される光増感剤の含有量は,材料の全質量に対して 1 質量%未満である。
校正結果	① 1.眼内レンズであって,それは以下を含む:
	② -マトリックス材料であって,疎水性アクリレート,アクリレートヒドロゲル,シリコーンゲル,シリコーンヒドロゲル,フルオロシリコーンアクリレート,ポリスチレンおよびポリメチルメタクリレート,ポリカーボネー

	ト,ポリシロキサン,またはその混合物から選択されるもの;
	③ -少なくとも 1 種の光増感剤であって,光力学型および/または光熱型から選択され,光増感剤が重合性基を含むもの;
	④ ここで,光増感剤はマトリックス材料と共役結合の方式で結合され,以下から選択される:
	⑤ -光増感剤がマトリックス材料の成型過程で重合に加わる;および/または
	⑥ -光増感剤が表面グラフト,表面修飾の方式でマトリックス材料表面に固定される;
	⑦ ここで材料の全重量に対して,含まれる光増感剤の含有量は 1 重量%未満である。

※考察の便宜上、各欄冒頭に①～⑦を付している。

本請求項はセミコロンと意図的改行によって 7 行に分割されている。機械和訳は、これら 7 行のうち④行目と⑤行目が訳抜けしている。④行目に相当する位置には「光活性剤および/または光活性剤からなる群から選択される少なくとも 1 種の光増感剤;」という訳文が出力されているが、これは③行目の内容が（不正確に）繰り返されたものであり、本来の④行目の内容「ここで,光増感剤はマトリックス材料と共役結合の方式で結合され,以下から選択される:」とは内容が大きく異なる。これを原文④行目が誤訳された結果と見るのは無理があり、④行目の訳抜けと、③行目の内容の重複出力（湧き出し）の併発と見なすのが妥当である[10]。⑤行目は、セミコロン後の「および／または」のみ出力されているが、「光敏剤在基体材料成型过程中参与聚合;（光増感剤がマトリックス材料の成型過程で重合に加わる;）」という本文全体が訳抜けしている。

本例の④行目、⑤行目は、文長は短く、特段の難訳語も存在しない。このため、翻訳の難易度は低いと見られる。それにもかかわらず、セミコロン区切り全体にわたる大規模な訳抜けが発生している。機械翻訳プラットフォームの機械和訳では、このように意図的改行やセミコロンで適度な翻訳処理単位（≒行）に区切られた請求項においても、翻訳処理単位全体が欠落するレベルの大規模な訳抜けが発生するケースが多数見られた。

その反面、調査対象全件を通じて、請求項全体の訳抜けは一切発生していない。短い請求項は一回の翻訳処理で全体が翻訳されるはずだが、その全体が訳抜けする（つまり請求項全体が欠落する）ケースは見られなかった。この状況から、上例のような大規模な訳抜

[10] このように訳抜け部分を埋め合わせる形で文中の別の箇所が重複出力される事象については、次項（第Ⅲ.4 項）で述べる。

けは、複数に区切られて翻訳される長大な請求項のみで発生する事象と見なせる。第II章の統計分析で、請求項が長文となるにつれ重大な訳抜けの発生が顕著に増加していた状況も、これと整合する。

なぜこのような訳抜けが発生するかは確言できないが、訳抜けしている箇所の内容からは、語句の翻訳難易度の問題とは考えにくい。むしろ、機械翻訳プラットフォームではセミコロン等で文を分割していない可能性や、長大な請求項を複数に区切って翻訳処理し、その結果を連結して翻訳文を組み立てるという一連の処理の過程で、翻訳処理単位の一部が正しく処理されなかったり、もしくは翻訳文が消失したりするなど、長大な入力文に対するエンジンの翻訳処理もしくはその前後処理の失敗である可能性が高い。

Ⅲ.4. 湧き出し

湧き出しとは、原文中に存在しない内容・文言が機械和訳中に出力される不備を指す。機械翻訳プラットフォームにおける湧き出しの典型パターンとして、（前項の事例でも見られたように）原文の一部が訳抜けし、これを埋め合わせるように文中の別の部分が重複して出力されるケースが多数見られた。

訳抜けと湧き出しの併発は、必要な情報が失われ、かつ余分な情報が与えられるため、請求項の内容理解に重大な悪影響を及ぼす。根本的な発生原因は文の一部が訳抜けしたことにあり、それが湧き出しを誘発していると考えられる。長大な請求項で発生し易い不備であることは確かだが、必ずしも難訳語を含む文や複雑な構造の文で発生するとは限らず、発生条件の特定は難しい。

なお、湧き出しのもう一つの典型パターンとして、頻出の特許用語・表現が文中の用語や文脈から連想される形で湧き出すケースが散見された。以下に例示する。

［事例Ⅲ.4-1：CN107404922B ｾｸｼｮﾝ A］

原　　文	17.包含通过根据前述权利要求中任一项所述的方法获得的乳基蛋白质水解物的组合物。
機械和訳	17.請求項 1～10 のいずれか一項に記載の方法により得られた乳ベースタンパク加水分解物を含む組成物。
校正結果	17.前記請求項のいずれか一項に記載の方法により得られた乳ベースタンパク加水分解物を含む組成物。

本事例では、原文が「前述权利要求中任一项所述（前記請求項のいずれか一項に記載の）」と、特定の請求項番号を記していないのに対し、機械和訳は「請求項 1～10 のいず

れか一項に記載の」と、原文に書かれていない請求項番号が湧き出している。類似の不備として、機械和訳の末尾に「請求項 1 に記載の方法」など、原文に存在せず、内容的にも無関係な文言が湧き出しているケースも散見された。

こうした湧き出しは、突発的ながら一定の頻度で発生している。訳抜けがなく、かつ原文中に存在しない文言が湧き出していることから、訳抜けを埋め合わせるタイプの湧き出しとは性質が異なる。湧き出す文言が請求項で多用されるものであることから、請求項由来の学習データの精度の問題である可能性がある。

機械翻訳プラットフォームの学習データには中日対応特許由来のコーパスが使用されている可能性が高い。しかしながら、対応特許であっても双方が必ずしも完璧な対訳文になっているとは限らない。特に特許請求の範囲は、請求項の数や順序が変わっていたり、記載内容や表現・様式が微妙に違ったりして、不正確な対訳文となりがちである。本項で挙げた二例は、湧き出している文言が請求項で多用されるものであることから、こうした不正確な対応関係の請求項に由来する学習データが原因である可能性がある。

Ⅲ.5. 数値の処理

機械翻訳プラットフォームの機械和訳には、原文中の数値の処理に関連する不備が目立った。特に重大なものとして、漢数字の処理に関する不備が挙げられる。以下、一例を示す。

［事例Ⅲ.5-1：CN109736014B ｾｸｼｮﾝ D］

原　　文	7.根据权利要求 6 所述的一种六边形 FRP 筋材网片的制造方法，其特征在于，所述限位块的材质为聚四氟乙烯、聚甲醛或硬质硅胶。
機械和訳	7.請求項 6 に記載の 6 角形 FRP 筋材メッシュの製造方法であって,その特徴は以下のとおりである:前記ストッパブロックの材質はポリテトラフルオロエチレン,ポリオキシメチレン又は硬質シリカゲルである。4。
校正結果	7.請求項 6 に記載の六角形 FRP リブメッシュの製造方法であって,その特徴は以下のとおりである:前記位置決めブロックの材質はポリテトラフルオロエチレン,ポリオキシメチレン又は硬質シリカゲルである。

本例の機械和訳は、文末に「4」と出力されている。原文中に「4」という算用数字は存在せず、「聚四氟乙烯」という化合物中に漢数字「四」が存在するのみである。同様のケースが頻発していることから、機械翻訳プラットフォームには、文中の漢数字から算用数字を生成して記憶する処理が組み込まれていると考えられる。

この処理は、本例の別箇所の「六边形」→「6 角形」など、漢数字部分をそのまま算用

数字に置換できるケースでは正しく機能する。だが、「聚四氟乙烯」のように漢数字が用語の一部として用いられ、算用数字に置換すべきでないケースでこの処理が働くと、生成された算用数字が行き場を失い、やむなく文末に出力されるものと推察される。

この現象は漢数字に準ずる漢字でも発生する。とりわけ「两（両)」という文字で顕著であり、例えば「两側」が「両側」と訳出される一方で、別途派生した「2」が文末に出力されるケースが多発していた。

さらに、漢数字から算用数字が派生したことで、より重大な不備を招く事態も散見された。下例は、漢数字からの派生によって原文よりも多くの算用数字が存在することとなった結果、余った数値が強制的に文脈に組み込まれたと考えられるケースである。

［事例Ⅲ.5-2：CN109044881B ｾｸｼｮﾝ A］

原　　文	4)投入非硫酸盐两性表面活性剂、非硫酸盐非离子表面活性剂，搅拌均匀，降温至 40-35℃；出料。
機械和訳	4)は非硫酸塩性界面活性剤、非硫酸塩系非イオン界面活性剤を投入し,均一に撹拌し,温度を 2℃まで下げて 40-35℃にする;。
校正結果	4)非硫酸塩両性界面活性剤,非硫酸塩非イオン界面活性剤を投入し,均一に撹拌し,40-35℃まで降温する;完成品を取り出す。

本例の機械和訳は、本来「40-35℃まで降温する」であるべき箇所に「温度を 2℃まで下げて 40-35℃にする」と不要な文言が湧き出している。原文中に算用数字「2」は存在せず、「两性」の「两」から「2」が派生したと考えられる。通常、こうしたケースでは不要な「2」が文末に出力されるが、本例では「2℃まで下げて」と不要な文言を伴って文脈に組み込まれるという、複雑かつ重大な不備となっている。

このように、漢数字から算用数字が派生することに起因する不備の事例は極めて多い。あらゆる文で発生していることから、特定の学習データが原因の現象とは考えにくく、システム上の数値処理が関与していると推察される。

なお、中国語の化合物名称には漢数字が使用されることが非常に多い。前掲の事例Ⅲ.5-1 も、化合物名称中の漢数字に起因した不備事例であった。この事例では化合物名称自体は正しく訳されていたが、化合物名称中の漢数字が算用数字に置換されたことによって化合物名称自体に誤訳が生じるケースも多数検出されている。一例を示す。

［事例Ⅲ.5-3：CN109562045B ｾｸｼｮﾝ A］

原　　文	4.根据权利要求 1 至 3 中任一项所述的组合物，其中所述组合物还包含由选自锌盐、亚锡盐和氯化十六烷基吡啶的口腔可接受的盐提供的口腔可接受的阳离子型活性剂。
機械和訳	4.請求項 1 から 3 のいずれか一項に記載の組成物であって,ここで前記組成物はさらに亜鉛塩、第一スズ塩及び第 16 塩化ジアルキルピリジンから選択される口腔に許容可能な塩により提供される口腔内に許容可能なカチオン性活性剤を含む。
校正結果	4.請求項 1 から 3 のいずれか一項に記載の組成物であって,ここで前記組成物は亜鉛塩,第一スズ塩及び塩化セチルピリジニウムから選択される口腔用として許容可能な塩により提供される口腔用として許容可能なカチオン活性剤をさらに含む。

本例では、化合物名「氯化十六烷基吡啶（塩化セチルピリジニウム)」が機械和訳では「第 16 塩化ジアルキルピリジン」と誤訳されている。語中の漢数字「十六」が「16」に置換され、名称に組み込まれた形である。

中国対応特許では「氯化十六烷基吡啶」は 214 文の使用例があり、うち 162 文で「塩化セチルピリジニウム」と訳されている。このため、学習データ中でも支配的な訳語であると考えられる。一方、中国対応特許に「第 16 塩化ジアルキルピリジン」という訳例は存在しない。

こうした状況にもかかわらず、本調査対象の請求項で「氯化十六烷基吡啶」が使用されている 13 件（本例を除く）のうち、正しく「塩化セチルピリジニウム」と翻訳されたのは 8 件のみであった。残る 5 件は本例と同様、「16」＋「塩化ジアルキルピリジン」と誤訳されている。他にも、化合物名に含まれる漢数字が算用数字に置換されたことで学習データの適訳語とマッチせず、誤訳されたと考えられるケースは非常に多かった。

Ⅲ.6. 請求項の記載様式に係る不備

機械翻訳プラットフォームの請求項の機械和訳における不備の中には、中国及び日本の請求項の記載様式に係るものも見られた。

Ⅲ.6.1. 請求項番号が本文の一部として扱われる不備

中国特許公報の請求項は、冒頭に「半角数字＋半角ピリオド」（例：「1.」「2.」）のフォーマットで請求項番号が記載されている。

通常、請求項番号は機械和訳にそのまま反映される。だが、一部の請求項では、請求項番号が本文の一部と誤認され、不要な文言を伴った形で本文に組み込まれていた。以下、典型的な事例を示す。

［事例Ⅲ.6.1-1：CN108180408B ｾｸｼｮﾝ F］

原　　文	1.一种中继镜头，其特征在于，所述中继镜头包括第一部分和第二部分；所述第一部分的第一入射面为与光路垂直的平面；所述第一部分的第一出射面为凸球面；所述第二部分位于所述第一部分的中心位置；所述第二部分的第二入射面为与所述光路垂直的平面；所述第二部分的第二出射面的曲率小于所述第一部分的第一出射面的曲率；所述第二出射面的最大截面尺寸，通过如下方式得出：计算不同的所述第二出射面的截面尺寸与照明效率及照明均匀性的对应关系；选择所述照明效率和照明均匀性中的任一项达到设定最低值时对应的尺寸，作为所述第二出射面的最大截面尺寸。
機械和訳	請求項１に記載の方法において,前記リレーレンズを第一部分と第二部分とを具備し,前記第一の部分の第一の入射面は,光路と垂直な平面であり,前記第一の部分の第一の出射面が凸球面である前記第二の部分は前記第一部分の中心位置に位置しており,前記第二の部分の第二の入射面が,前記光路と垂直な平面であり,前記第二の部分の第二の出射面の曲率が前記第一の部分の第一の出射面の曲率よりも小さく,前記第二の出射面の最大断面寸法は,前記第二出射面の異なる断面サイズと光効率と照明均一度との対応関係を算出し,前記照明効率および照度の均一性のいずれかに設定され,最も低い値に対応する大きさに到達するように選択し,前記第二の出射面の最大断面寸法であることを特徴とするリレーレンズである。
校正結果	1.リレーレンズであって,その特徴は以下のとおりである:前記リレーレンズは第一部分と第二部分を含み,前記第一部分の第一入射面は光路と垂直な平面であり,前記第一部分の第一出射面は凸球面である;前記第二部分は前記第一部分の中心位置に位置し,前記第二部分の第二入射面は前記光路と垂直な平面であり,前記第二部分の第二出射面の曲率は前記第一部分の

	第一出射面の曲率よりも小さい;前記第二出射面の最大断面寸法は以下の方式で取得される:前記第二出射面の断面サイズごとの照明効率と照明均一性との対応関係を算出する;前記照明効率および照明均一性のいずれか一つが設定された最低値に達した時の対応する寸法を選択し,前記第二出射面の最大断面寸法とする。

上例では、原文冒頭の請求項番号「1.」が「請求項 1 に記載の方法であって」という誤った解釈に基づいて機械和訳の本文に組み込まれている。「図 1 は,」と誤認されるケースもこれに次いで多く、他にも多数のバリエーションが見られた。

請求項番号が本文中に組み込まれる不備は、請求項 2 以降でも同様に発生している。下表に請求項 1～5 における発生頻度を示す。

	全請求項		不備発生請求項		
	請求項数	平均文長	請求項数	発生率	平均文長
請求項 1	100,000	460.2 文字	4,455	4.5%	519.1 文字
請求項 2	95,777	107.0 文字	274	0.3%	320.1 文字
請求項 3	91,953	102.6 文字	247	0.3%	302.9 文字
請求項 4	86,914	104.7 文字	274	0.3%	318.5 文字
請求項 5	80,575	106.0 文字	278	0.3%	316.1 文字

不備は各請求項で発生しているが、請求項 1 の発生率が特に高く、請求項 2 以降では長大な請求項に偏って発生している。このことから、この不備は長大な請求項で発生する傾向が強いことがわかる。請求項 1 での発生率が高いのは、請求項 1 が独立請求項であり、概して長大なためと考えられる。

請求項番号が原文と無関係な文言とともに翻訳文に組み込まれるというこの不備の状況は、第Ⅲ.5 項の事例Ⅲ.5-2 で見られた数値処理の不備、すなわち「両」から派生した「2」が「2℃まで下げて」という無関係な文言を伴って翻訳文に組み込まれた状況と似ている。このことから、機械翻訳プラットフォームのエンジンは、請求項番号も本文中の数値と同様に扱っていることが窺える。つまり、この不備の発生原因は、入力文が長大なことによる翻訳処理精度の低下と、数値処理上の問題とが複合していると考えられる。

Ⅲ.6.2. 発明構成要素の訳ゆれ

訳ゆれとは、同一の用語が、一請求項内や一案件内において複数の訳語に訳されている状態を指す。一例を示す。

［事例Ⅲ.6.2-1：CN107074359B ｾｸｼｮﾝ B］

原　　文	2.根据权利要求 1 所述的航空器，其中所述止挡构件被配置成使所述推进器通过以下动作中的至少一者停止：从所述马达移除电流；使至所述马达的所述电流的极性逆转；将止挡杆部署至所述马达的转子中；或使所述推进器从所述马达脱离。
機械和訳	2.請求項 1 に記載の航空機であって,ここで前記ストッパ部材は前記スラスタが以下の動作のうちの少なくとも 1 つによって停止するように配置される:前記モータから電流を除去する;前記モータへの前記電流の極性を反転させる;ストッパロッドを前記モータの回転子に配置する;又は前記推進器を前記モータから離脱させる。
校正結果	2.請求項 1 に記載の航空機であって,ここで前記ストッパ部材は前記プロペラが以下の動作のうちの少なくとも一つによって停止するように配置される:前記モータから電流を除去する;前記モータへの前記電流の極性を反転させる;ストッパロッドを前記モータの回転子に配置する;又は前記プロペラを前記モータから離脱させる。

本件は、原文中に二度出現する「推进器」が、機械翻訳では「スラスタ」と「推進器」に訳ゆれしている。このため、一読して両者が同じ構成要素を指すとは認識し難い。

本件は自動航空機（ドローン）の発明であり、「推进器」は「プロペラ」と訳すのが適切である。だが、本件の全請求項で「推进器」が 37 回出現するうち、機械和訳で「プロペラ」と訳されたのは 4 回のみであり、あとは「スラスタ」が 13 回、「推進器」が 10 回、「プッシャ」が 6 回、「推進装置」が 2 回、「インペラ」が 1 回と、一案件中で 6 種の訳語に訳ゆれしていた（残る 1 回は訳抜け）。

ニューラル機械翻訳で訳ゆれが発生する原因は、学習データ中に一つの用語に対して複数の訳語候補が存在することにある。ニューラル機械翻訳は、同一の用語であっても、周辺語句の状況等に応じ、学習データ中から最も適切な（≒ベクトル値が近似する）訳語をその都度選択する。これにより技術分野や文脈に応じた適切な訳し分けが可能となるが、その反面、同一の用語に対して異なる訳語が採用される可能性も常にはらむ。

用語の訳ゆれは、例えば動詞や一般用語であれば、内容理解を阻害しないことが多い。

だが、技術用語、とりわけ特許請求の範囲における発明の構成要素は、全請求項で統一名称で記載されるのが共通認識となっており、訳ゆれが発生すると、読み手の読解を妨げ、場合によってはそれぞれが別個の要素と誤認されて、発明の内容理解に支障をきたす。本事例はその典型例である。

用語の訳ゆれは、学習データからその都度最適の訳語を採用するニューラル機械翻訳方式においては宿命的な現象といえる。現状のニューラル機械翻訳で特定の語の訳語を統一するには、学習データから他の訳語を全て排除する必要があり、現実的ではない。

このため、現状では、請求項内や請求項間での用語の訳ゆれは解決困難である。ただし、ニューラル機械翻訳分野では、この課題に対処すべく、前後の文の文脈を考慮して訳語の統一を図る手法の研究が進められている。将来的には、請求項単位や案件単位で訳語情報を共有し、訳語の統一が達成されることが期待される。

Ⅳ. 本調査のまとめ

本調査により把握した、機械翻訳プラットフォームにおける中国特許請求項の機械和訳の翻訳品質及び不備の傾向と、さらなる精度向上のための提言を本章にまとめる。

Ⅳ. 1. 機械翻訳プラットフォームの現状の翻訳品質について

本調査では、中国特許公報 10 万件の全請求項 949,882 項に対する校正作業の記録に基づき、現状の機械翻訳プラットフォームにおける機械和訳の不備の状況を調査分析した。

第 II 章に示した統計的分析では、校正者によって何らかの修正が加えられた請求項は全体の 86%を超えたものの、発明の内容理解を大きく阻害する「重大な不備」の発生率は 21.4%に抑えられていることがわかった。ただし、独立請求項と従属請求項に区別した集計では、独立請求項において重大な不備の発生率が 67.1%と非常に高いことも判明した。従属請求項の重大不備発生率は 13.4%と低かったが、従属請求項は引用した独立請求項の内容を踏まえて理解すべきものであり、したがって実質的な翻訳精度は独立請求項を基準に判断するのが妥当である。つまり、現状の機械翻訳プラットフォームにおける最大の課題は、67.1%と高頻度で発生している独立請求項の重大不備を減少させることにある。

独立請求項で重大不備が多発する最大の要因は、その極端に長大な文長にあると考えられる。請求項は一文で記述するのが通例であるため、発明の全体像が詳細に記載される独立請求項は、一般的な文に比べて極端な長文となる傾向が強い。本調査の対象の 10 万件

においても、従属請求項の平均文長が 88.7 文字であったのに対し、独立請求項の平均文長は 400 文字を超えていた（420.4 文字）。

機械翻訳は一文単位の処理が基本であり、その翻訳精度は原則として文長が長くなるほど低下していく。事実、本調査の統計的分析でも、請求項の文長と重大不備の発生率との間に明瞭な相関性が確認された（⇒第Ⅱ.3 項）。具体的には、「重大な誤訳」が文長に比例してコンスタントに増加していくことに加え、文長が 150 文字を超えると、それまで少数であった「重大な訳抜け」と「重大な湧き出し」も急増する傾向が見られた。そして、この傾向は基本的には独立請求項、従属請求項の双方に共通した。

長大な請求項では、意図的改行やセミコロンが多用される。機械翻訳プラットフォームでは、機械和訳の状況から、他の多くの機械翻訳システムと同様、これらを文の分割位置として利用していると推測される。つまり、長大な請求項も意図的改行／セミコロンで適宜分割され、実際の機械翻訳処理の単位は適度な文長に抑えられていると考えられる。一文を分割して翻訳する手法は、分割された各部分間の係り受け関係が失われるデメリットがあるが、請求項の場合、例えばステップや構成要素ごとに意図的改行・セミコロンで区切られ、内容的にも文法上も各区切り単位で文として独立していることが多い。このため、分割して翻訳し、単純に連結する方法でも、翻訳文が破綻なく成立することが多いと予測された。

事実、統計的分析でも、同程度の長文であれば、適度に意図的改行やセミコロンが挿入されたもののほうが重大不備の発生率は低かった（⇒第Ⅱ.4 項）。だが、これを踏まえても長大な請求項における不備発生率は依然として高く、重大な訳抜けや湧き出しが顕著に増加する傾向も、改行・セミコロンの多寡にかかわらず見られた。機械翻訳プラットフォームではセミコロン等で必ずしも文分割を行わない可能性も考えられるが、多数の不備事例の状況から総合的に判断すると、長文の処理には文分割のみでは解決されない課題がある可能性のほうが高い。いずれにせよ、機械翻訳プラットフォームには、長文の翻訳品質に大きな改善の余地がある。

統計的分析ではさらに、技術分野（IPC セクション）別、出願人の国籍別、法人／個人の別などの観点でも不備の発生頻度を比較調査した。その結果、各観点とも文長差に準拠した不備発生率となっており（⇒第Ⅱ.5～Ⅱ.6 項）、請求項の文長が機械翻訳プラットフォームの翻訳精度を決定する最大のファクターであることが改めて示された。

Ⅳ.2. 機械翻訳プラットフォームの特徴的な不備と対処策について

本調査ではさらに、現行の機械翻訳プラットフォームに見られる特徴的な不備を収集し、類型別に分析した（⇒第Ⅲ章）。これを踏まえ、中国特許請求項の機械和訳において重大な不備をもたらす主な要因、つまり最優先の課題には以下の四点が挙げられる。

- 学習データが最新の技術用語に対応できていない
- 前後処理を含めた長文の処理精度が不十分である
- 数字（漢数字を含む）の処理方式に改善の余地がある
- 請求項冒頭の請求項番号が本文に組み込まれて誤解釈される

以下、上記各点について、具体的な状況と、採り得る改善策を提言する。

Ⅳ.2.1. 学習データが最新の技術用語に対応できていない

特許文献用のニューラル翻訳システムでは、対応特許由来のコーパスが学習データに用いられることが多い。機械翻訳プラットフォームも、学習データの内容は非公開ながら、主に対応特許コーパスを使用している可能性が高い。

ニューラル機械翻訳では、学習データ中に存在しない語彙は正しく翻訳することができない。請求項には非常にマイナーな技術用語や出願人の造語が使われることもあり、対応特許由来の学習データでこれら全ての語彙をカバーすることは不可能である。このため、ある程度の誤訳、不適訳の発生は事実上避けられない。

しかしながら、本調査の結果、現行の機械翻訳プラットフォームでは、「ディープラーニング（深度学习）」や「ビッグデータ（大数据）」、「充電スタンド（充电桩）」など近年になり一般的に使われ出したメジャーな技術用語も常時誤訳されていることが判明した（⇒第Ⅲ.2.1 項）。これらの語句は中日対応特許でも既に多用されており、他種のニューラル機械翻訳エンジンでは概ね正しく翻訳される。

このことから、機械翻訳プラットフォームの機械翻訳エンジンは、他のエンジンに比べてアップデートが遅れている可能性が高い。特許文献は最先端の技術が記載される文書であることから、定期的に最新の対応特許由来のコーパスを追加学習させ、語彙を増強した機械翻訳エンジンを用いることが有効と考える。学習データの増強は、マイナーな技術用語のカバー範囲を広げることにもつながる。

Ⅳ.2.2. 前後処理を含めた長文の処理精度が不十分である

機械翻訳プラットフォームでは、特に文長 150 文字を超える請求項において、訳抜けや湧き出しなどの重大な不備が多発する傾向が顕著であった。この傾向は、独立／従属請求項、技術分野、出願人の国籍、意図的改行／セミコロンの多寡などにかかわらず一律に見られ、現行の機械翻訳プラットフォームが、特許請求項の「一文で記載されるため極端に長大になる傾向が強い」という特徴に十分に対応できていないことを示している。独立請求項の平均文長は 400 文字を超えており、状況の改善が求められる。

長文の翻訳精度改善の一手段として、新方式のニューラル機械翻訳である Transformer 方式の採用が推奨される。この方式は、従来のニューラル機械翻訳の弱点であった、距離の隔たった語間の関係性の把握に優れる。このため、特に長文の翻訳精度改善に有効であり、特許請求項の翻訳用途への適性が高いと考える。

ただし、長文の翻訳精度改善には、翻訳の前後処理の精度も等しく重要である。機械翻訳プラットフォームでは、長大な請求項は意図的改行やセミコロンで適度な文長に区切って翻訳されていると推測されるが、こうして適度な文長に分割して翻訳された請求項においても、大規模な訳抜けや湧き出し等の不備は依然多発しており、顕著な改善は見られない（⇒第Ⅱ.4 項、第Ⅲ.3 項）。類型ごとの文長別発生率（⇒図Ⅱ.3-1～2）を見ても文長 150 文字以下の短い請求項では重大な訳抜けや湧き出しがほとんど発生していないことから、長文を区切って順番に機械翻訳エンジンに供する前処理、又は順次翻訳された各文を一文に連結して出力する後処理においても改善の余地がある可能性が高い。

事実、Transformer 方式を採用している他種エンジンの中には機械翻訳プラットフォームと同様の不備が頻繁に発生しているものもあり、長文の翻訳精度向上は Transformer 方式の採用のみで達成されるとは限らないことがわかる。機械翻訳プラットフォームにおいて長大な請求項の精度向上を果たすには、Transformer 方式の採用と、長文に対する文分割・再連結などの前後処理の精度向上の双方が等しく重要であると考える。

Ⅳ.2.3. 数字（漢数字を含む）の処理方式に改善の余地がある

機械翻訳プラットフォームでは、数字・漢数字に関連する特徴的な不備が多く見られた（⇒第Ⅲ.5 項）。なかでも悪影響が大きなものとして、原文中の漢数字が算用数字に置換された結果、漢数字を含む化合物名称などの技術用語が正しく扱われず、学習データ中に訳例が存在するにもかかわらず誤訳されてしまう不備や、漢数字から派生した算用数字が不要な文言を伴って翻訳文に組み込まれる不備などが挙げられる。

中日対応特許コーパス（すなわち学習データ）には、原文は漢数字、翻訳文では算用数字が用いられる対訳文も存在するが、他種エンジンでは同様の不備は見られない。したがって学習データが原因とは考えにくく、機械翻訳プラットフォームに搭載されたエンジンの数字・漢数字処理に起因した不備である可能性が高い。

特許請求項において数値は重要な情報である。加えて、漢数字は中国語では化合物名称をはじめ様々な語句に用いられており、数値関連の不備の影響範囲はきわめて広い。これらの不備が機械翻訳プラットフォームのエンジンに起因するものであるならば、その更新が必要となる。学習データのアップデート、長文の処理精度向上とともに、最優先で対処すべき課題であろう。

Ⅳ.2.4. 請求項冒頭の請求項番号が本文に組み込まれて誤解釈される

第Ⅲ章で「請求項の記載様式に係る不備」として挙げた「請求項番号が本文の一部として扱われる不備（⇒第Ⅲ.6.1 項）」も、請求項 1 全件の 4.5%で発生するなど発生頻度が高く、かつ請求項番号が誤った解釈で不必要な文言を伴って本文に組み込まれることで請求項全体の翻訳精度に大きな悪影響を及ぼすため、優先的に対処が必要な課題である。

この不備は長大な請求項に偏って発生していることから、第Ⅳ.2.2 項で述べた長文の処理精度という課題と関連している可能性が高い。かつ、数値が無関係な文言を伴って翻訳文に組み込まれるという点において、第Ⅳ.2.3 項で述べた数値処理関連の不備にも該当する。このため、これらへの対処により同時に改善される可能性もある。または、請求項番号を前処理であらかじめ本文と切り離し、機械翻訳処理を迂回させて、後処理で翻訳文に戻すという手法も有効であろう。

Ⅳ.3. 現状では改善困難な不備

本調査で検出された不備の類型の中には、他種のエンジンでも等しく発生し、現状のニューラル機械翻訳技術では改善が難しいと考えられるものも存在した。以下、主要な二種について述べる。

Ⅳ.3.1. マイナーな技術用語や出願人による造語の誤訳

「学習データ中に存在しない誤訳（⇒第Ⅲ.2.1 項）」の多くを占めたマイナーな技術用語や出願人の造語は、中日対応特許にも使用例が存在しない場合が多い。こうした語を正しく翻訳させるためには、対訳データを自ら作成して学習に供する必要がある。だが、これを行うには、まずは機械和訳を解析して誤訳されている語句を収集する必要があり、さらに対訳データを人手翻訳で作成しなければならない。これには多大なコストを要し、かつ、こうした語は無限に発生するため際限のない対応となる。

マイナーな用語や出願人の造語はごく少数の案件のみで用いられる傾向にあるため、対応しても全体的な翻訳精度の改善効果は小さい。現状ではこうした語彙を網羅的にカバーできる決定的な打開策はなく、コストパフォーマンスの観点からの割り切りが必要であろう。

Ⅳ.3.2. 同一の技術用語の訳ゆれ

特許請求項では、同一の技術用語が発明の構成要素として繰り返し言及される傾向が強い。同一案件中の多数の請求項に共通の技術用語が用いられることも通常である。

一方、ニューラル機械翻訳は、各用語の訳語を学習データに基づいて都度決定している。このため、同じ用語であっても分野や文脈に応じて自動的に訳し分けることができるが、その反面、同一案件や同一請求項で同じ語が異なる訳語に「訳ゆれ」する可能性も生じる。事実、本調査では、このような発明構成要素の訳ゆれが非常に多く検出された（⇒第Ⅲ.6.2 項）。

現状、こうした訳ゆれはニューラル機械翻訳の方法自体が内包する宿命的なものといえ、即効性のある対処策は見当たらない。ただし、ニューラル機械翻訳分野では現在、この課題への対処策として、前後の文の文脈を考慮して（つまり、先行文で用いた訳語情報を保存するなどして）訳語の統一を図る手法の研究が進められている。将来的には、請求項単位や案件単位で訳語情報を共有することで、訳ゆれの解消が進むものと期待される。

調 査 報 告

1. 本事業の概要

本報告書は、2022 年 4 月から 2023 年 3 月まで 1 年にわたり実施した「中国特許の請求項及び中国特許審決の機械翻訳に関する調査事業（以下「本事業」という。)」の調査報告書である。本章で、本事業の概要についてまとめる。

1.1. 本事業の背景・目的

中国において日本国企業が特許を効果的に取得し、特許紛争に対して適切な対処をするためには、特許の権利範囲及び審判事件の傾向を把握する必要がある。すなわち中国特許の請求項及び中国特許審決の照会環境整備が急務である。

特許庁は、機械翻訳プラットフォームにより中国特許公報全文の機械和訳を作成し、特許情報プラットフォーム（J-PlatPat）を通じて提供している[11]。また、中国特許審決の機械和訳に特化したニューラル機械翻訳エンジンの学習用対訳コーパスの作成事業を実施した[12]。

中国特許審決については、上述した事業で作成された学習用対訳コーパスを活用した、高品質な全文機械和訳を速やかに作成して提供することが望ましい。

他方、中国特許の権利範囲の把握に関しては、権利への抵触防止等の観点において、その内容を正確に理解する必要がある。しかし、上述した現在提供中の中国特許公報全文の機械和訳にも、一定の誤訳が混入している。したがって、権利範囲を定める請求項の機械和訳については、調査・分析をして誤訳傾向を明らかにすることで品質向上に取り組むことが望ましい。誤訳傾向の調査・分析は、機械和訳に誤訳が含まれているかどうかを解析し、その誤訳がどのような校正を行うことで解決する誤訳であるか、その傾向を調べることによって行うことができる。調査・分析の過程で作成される人手校正文は、高品質な和訳として提供することが望ましい。

本事業では、上記した学習用対訳コーパスを活用した中国特許審決の高品質な全文機械和訳を作成した。具体的には、特許庁の作成した中国特許審決由来の学習用対訳コーパスを一般財団法人 日本特許情報機構（以下、「当財団」という。）保有のニューラル機械翻訳エンジンに学習させ、そのエンジンを使用して、6 万件の中国特許審決の全文機械和訳を作成

---

[11] 2015 年 1 月から、中韓文献翻訳・検索システムにより提供開始。2020 年 5 月以降は機械翻訳プラットフォームにより提供。

[12] 中国審決情報の機械翻訳に関する解析及びデータ作成事業(平成 30 年度から令和 3 年度)において、200 万コーパスを作成した。「みんなの自動翻訳@Textra®」におけるニューラル機械翻訳エンジンへ学習させることで、高い学習効果が確認されている。

した。

中国特許の請求項の機械和訳については、機械翻訳プラットフォームで提供されている10 万件を対象に、人手による校正の要否及び必要となる校正の程度に基づいて誤訳の有無とその内容を解析し、誤訳傾向を調査・分析した。これにより、権利範囲を定める中国特許の請求項のより正確な機械和訳を提供する上で、現在の機械翻訳プラットフォームが抱える問題の傾向（誤訳、訳抜け、湧き出しが発生しやすい傾向）を明らかにした。また、上記解析に伴って作成される中国特許の請求項の機械和訳の人手校正文を、高品質な和訳として提供する。これらを通じて、国内企業の中国における円滑な事業実施の支援を目指す。

1.2. 調査分析の内容

本事業では、中国特許公報（登録公報）10 万件の「特許請求の範囲」全文について、特許庁が機械翻訳プラットフォームで提供している機械和訳を人手で校正し、発明を正確に理解できる品質の翻訳文を作成した。その上で、各案件に対して実施した校正の内容を解析し、現状の機械和訳における誤訳の発生頻度や傾向・特徴を統計的に把握するとともに、現状の機械和訳の特徴的な誤訳・訳抜け・湧き出しについて調査し、実例に基づきその発生原因や解決策を分析した。本報告書で、その調査分析結果を報告する。

1.3. 調査対象 10 万件の選定基準

本調査の対象、すなわち人手校正の対象とする中国特許公報 10 万件は、以下の基準で選定した。

① 特許庁の提供する機械和訳が、最新の翻訳エンジンで作成されたもの

本調査は、特許庁が機械翻訳プラットフォームで提供している中国特許公報の機械和訳の品質改善を図ることを目的としている。このため、調査対象とする中国特許公報は全て、現行のニューラル機械翻訳エンジン[13]が用いられた 2020 年 3 月以降発行の公報[14]から選定した。

② 各 IPC セクション（A～H）から均等に選定

技術分野によって機械翻訳精度に偏りが生じていたり、特定の分野のみに特徴的な不備が発生している可能性が考えられる。このため、対象とする中国特許公報は、筆頭 IPC のセクション（A～H）別に 12,500 件ずつ均等に選定した。

---

[13] https://www.jpo.go.jp/support/j_platpat/kaizen20200518.html

[14] 2020 年 5 月の訳質改善後に機械翻訳された 2020 年 3 月発行分から、2022 年 4 月発行分まで。

③ 原文の「特許請求の範囲」の平均文字数

本事業の調達仕様に則り、対象とする各公報の「特許請求の範囲」全文、すなわち一公報中の全請求項の中国語原文の文字数が平均 1,300 文字[15]となるように調整した。さらに、IPC セクションごと（上記②）の案件集合においても平均文字数 1,300 文字を維持し、各集合間の条件を均一とした。

④ イメージタグ、テーブルタグを含む案件の除外

特許公報では、テキストで表現できないような複雑な数式などにイメージデータを使用することがある。イメージデータは機械翻訳では処理できないため、機械和訳に出力されない。テキスト化も困難であるため、人手校正で機械和訳に反映させることも難しい。

また、特許公報では、データを表形式で表示するためにテーブルタグが使用される場合もある。テーブルタグも機械和訳には反映されず、表中の内容のみが列挙されることとなる。かつ、表中に記載される内容は単なる数値等であることが多く、原文をそのまま出力すればよいため、人手校正が不要となることが大半である。

これらの理由から、イメージタグやテーブルタグを含む請求項は、人手校正の対象としても、機械和訳の誤訳調査の対象としても不向きであると判断し、対象から除外した。

⑤ 中国特許審決の対象となった案件

第 1.1 項に記したとおり、本事業では、その一環として、中国特許審決 6 万件の全文機械和訳を作成した。これら 6 万件の中国特許審決に対応する中国特許公報が存在する場合は、条件①、④に抵触するものを除き、全件を対象に選定した。その結果、8,022 件の審決対応公報が選定された[16]。

なお、中国特許公報 10 万件における審決対応の状況は、「対応審決リスト」として納入した。

1.4. 調査分析の手順

---

[15] 原文中の上付・下付文字タグ（<sup>、<sub>）は校正対象の機械和訳には反映されないため、案件選定時の文字数カウントからは除外した。ただし、機械翻訳の入力文という観点においては文長に含めるべきものであるため、本調査報告書の第 2 章以降に示す原文の文長ではこれらのタグもカウントに含めている。

[16] 6 万件の大部分（52,451 件）は復審審決であり、再審査を経て特許登録されるものはこのうちの一部である。かつ、特許公報の発行範囲を 2020 年 3 月以降の 2 年間に限定したため（⇒前ページ①参照）、再審査となっても事業実施時点で特許登録まで進んでいないものや、この範囲外のものは対象外となる。

調査分析は、以下の 2 つの手法により行った。

1.4.1. 機械翻訳傾向の統計的な調査・分析

調査対象の中国特許公報 10 万件の全請求項について、校正作業前後の和訳の内容を機械的に比較し、変化の有無（すなわち校正の有無）を取得した。さらに、校正時に重大な誤訳・訳抜け・湧き出しが検出された請求項に対しては、校正者自身がチェックマークを付し、発明の内容理解への悪影響が大きい重大な不備と、その他の比較的軽度な不備とを区別して集計できるようにした。

これら校正内容の記録を独立／従属請求項別、原文文長別、技術分野別など多角的な観点で集計し、機械翻訳プラットフォームにおける誤訳の発生傾向を統計的に調査・分析した。統計的調査・分析の結果は、第 2 章にまとめる。

1.4.2. 現状の機械和訳における特徴的な不備の調査・分析

校正作業を通じ、現状の機械和訳における特徴的な不備について、多数の機械和訳に共通して発生しているものや、内容理解への悪影響の大きいものを中心に実例を収集し、分類した。そのうえで、発生条件や発生原因の考察、他種のニューラル機械翻訳エンジンでの発生の有無の調査、これらを踏まえた改善可能性などについて考察した。特徴的な不備の調査・分析の結果は、第 3 章にまとめる。

2. 機械翻訳傾向の統計的な調査・分析結果

本事業で校正した中国特許公報 10 万件の「特許請求の範囲」の全請求項、949,882 項について、校正の状況を記録・集計した。本章に、その統計的調査・分析の結果を記す。

2.1. 校正状況の記録

校正の有無は、校正前後の機械和訳を機械的に比較して判定することができる。だが、機械和訳の不備には、誤解につながる重大なものもあれば、日本語として不自然な表現の修正、用語や体裁の統一など軽微なものもある。不備の種類にも、不適な訳語による「用語の誤訳」、原文の内容が訳出されない「訳抜け」、原文に無い内容が訳出される「湧き出し」などの類型が存在する。こうした不備の程度や類型は、機械的に判定することができない。

そこで、本調査では、校正の有無等の機械的判定による記録に加え、機械和訳に生じた「重大な誤訳」、「重大な訳抜け」「重大な湧き出し」を校正者が判定して記録することとした。不備が重大か否かの判定基準は、「原文の内容（特に技術内容）の正しい理解を大きく妨げるレベルの不備であるか否か」とした[17]。これらの記録を集計して、誤解に繋がる重大な不備の発生率や、不備の類型別の発生頻度を調査した。

なお、機械和訳が大きく崩れている、大規模な欠落があるなど、校正作業が困難な請求項に対しては、校正者からの申請により、当財団の機械翻訳システム[18]による再機械翻訳結果を提供した。再機械翻訳が使用された請求項は、当初の機械和訳に大規模な不備が発生していたと見なせる。このため、再機械翻訳の使用の有無も集計項目に加えた。

2.2. 重大な不備の判定精度の改善

機械和訳上の不備（誤訳、訳抜け、湧き出し）の重大度は物理的に計測できず、判定する者、すなわち校正者の主観に委ねる部分が大きい。本事業では 130 名の校正者を起用したが、本調査の実施の過程で、各人の重大不備の判定基準に個人差が大きいことが判明した[19]。このため、各人の判定内容をサンプルチェックするとともに、個々人の平均判定率[20]を算出して、全員の判定水準を揃える取り組みを行った。その結果、下図に示すとおり、事業前期に見られた平均判定率のばらつきが事業後期では顕著に改善された。具体的に

---

[17] 添付資料 01『「重大な誤訳」「重大な訳抜け」「重大な湧き出し」の判定基準』を参照されたい。

[18] Japio-AI 翻訳システム。特許文献専用に構築したニューラル機械翻訳エンジンを使用している。

[19] 不備を重大と見なすか否かに個人差が生じたものであり、校正作業自体の品質差は生じていない。次項（第 2.3 項）に示す「校正作業が発生した請求項」の比率も、事業全期間と事業後期分とがほぼ同値となっている。

[20] 校正者ごとに、担当した全請求項中で重大な不備と判定した請求項の比率を算出した。

は、前期分の標準偏差が 8.98%であったのに対し、後期分では 4.04%に良化している。なお、この比較は、対象案件 10 万件を、判定率の個人差が大きかった前期分 50,191 件（487,515 項）と、個人差が十分に改善された後期分 49,809 件（462,367 項）とに区分して行っている[21]。

図 2.2-1. 校正者別 重大不備チェック付与率分布（事業前期）

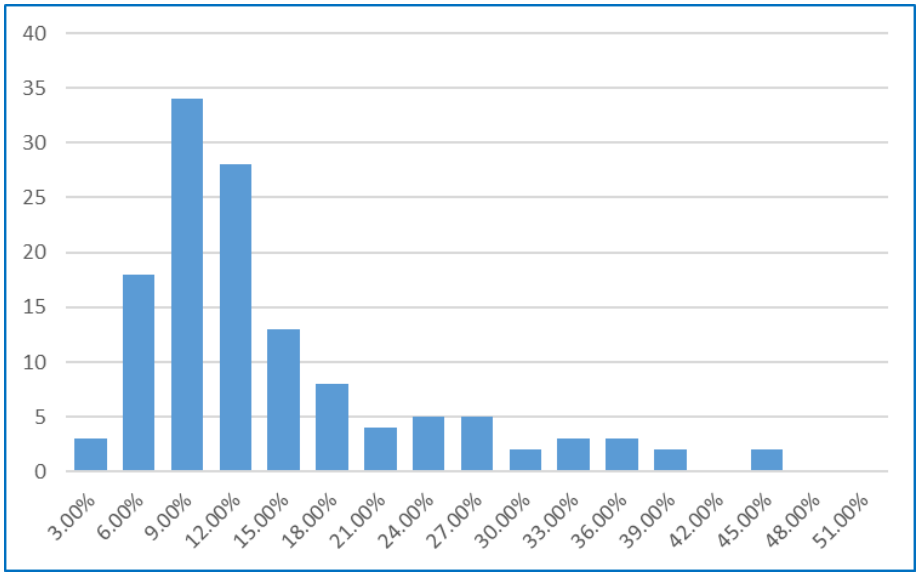
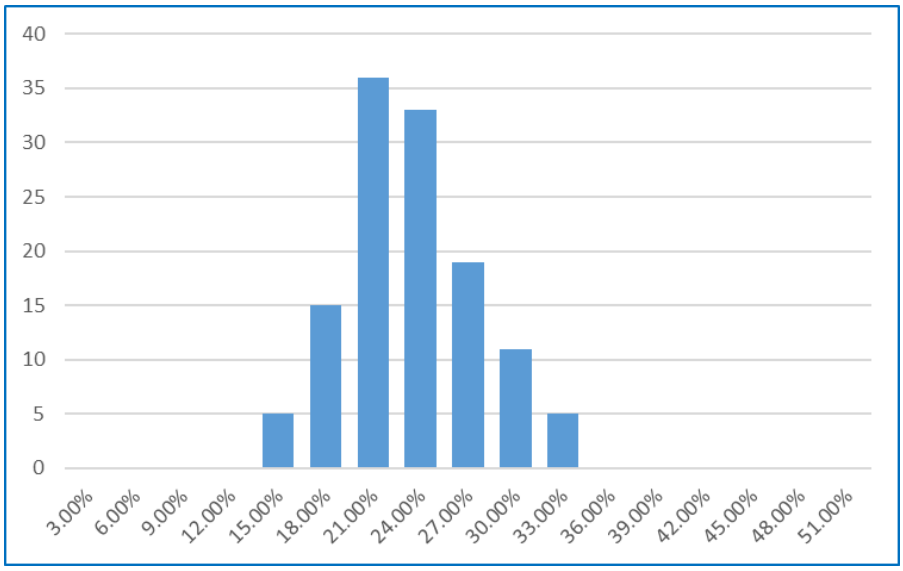


図 2.2-2. 校正者別 重大不備チェック付与率分布（事業後期）



[21] 本事業では 10 万件の校正作業を 20 のロットに分割して順次行った。事業後期分はこのうちの 12～20 ロットを指す。

こうした経緯から、重大な不備の発生頻度は、事業後期に校正された請求項のみを集計対象とすることで、より正確な現状把握になると判断される。ついては、本章に示す各項目の統計分析では、全請求項（949,882 項）の集計結果と事業後期分（462,367 項）の集計結果を併記するとともに、重大な不備に関する分析は原則として事業後期分の集計値に基づいて行うこととする。以降の表において、重大な不備に関しては、全請求項の集計結果にグレーで色付けしている。

2.3. 全請求項の集計結果と分析

本事業で校正を実施した中国特許公報 10 万件の全請求項、949,882 項における校正の頻度と、重大な不備の発生頻度を集計した結果を示す。あわせて、重大な不備の判定精度が向上した事業後期分（公報 49,809 件、請求項 462,367 項）のみの集計結果も併記する。表中各項目とも、上列が全期間、下列が後期分のみの集計値である。

表 2.3-1. 全期間及び事業後期分の校正と重大不備の発生率

請求項数		全	949,882 項
		後	462,367 項
一請求項あたりの平均文長		全	138.4 文字
		後	139.9 文字
一請求項あたりの平均改行数		全	2.1 回
		後	2.1 回
校正作業が発生した請求項		全	820,628 項（86.4%）
		後	400,517 項（86.6%）
重大な不備を含む請求項		全	181,563 項（19.1%）
		後	98,953 項（21.4%）
	・重大な誤訳を含む	全	137,032 項（14.4%）
		後	75,472 項（16.3%）
	・重大な訳抜けを含む	全	68,270 項（7.2%）
		後	37,308 項（8.1%）
	・重大な湧き出しを含む	全	41,899 項（4.4%）
		後	19,577 項（4.2%）
	・再機械翻訳を使用	全	5,554 項（0.6%）
		後	1,862 項（0.4%）

上表のとおり、本事業の校正作業では、全請求項の 86.4%に何らかの校正が加えられた。事業後期分のみの集計でも 86.6%とほぼ同値となった。原文の平均文長や平均改行数も同水準であり、全期間を通じて請求項の機械翻訳品質や物理的条件に偏りがないことが

わかる。

一方、重大な不備を含む請求項の比率は、全期間では 19.1%であったのに対し、事業後期分では 21.4%に増加した。事業前期のみの比率は 16.9%であり、前期と後期とで 4.5 ポイントの差が生じている。この差異は、事業開始当初の各人の判定に個人差が大きかったことに起因しており（⇒第 2.2 項)、各人の判定基準が統一された事業後期の数値のほうが信頼性が高いと考えられる。

重大な不備は類型ごとに記録している。事業後期分の重大不備を類型別に集計すると、「重大な誤訳」が 16.3%、「重大な訳抜け」が 8.1%、「重大な湧き出し」が 4.2%となった。また、再機械翻訳が使用された請求項も 0.4%存在した[22]。

上掲の集計結果から、機械翻訳プラットフォームの機械和訳では、大半（86%強）の請求項に何らかの校正すべき箇所が存在するが、内容の正確な理解を妨げる重大な不備を含むものは全体の 21%程度にとどまることが明らかとなった。

ただし、これらの集計値は、独立請求項と従属請求項、請求項ごとの文長や技術分野などの違いを区別せず、全ての請求項を一律に集計したものである。機械翻訳プラットフォームにおける中国特許公報の翻訳品質の傾向をより詳細に分析するためには、こうした請求項の特徴別に集計し、分析する必要がある。次項（第 2.4 項）より、これらの各観点からの統計分析を行う。

[22] 重大な不備の発生率は、「重大な誤訳」「重大な訳抜け」「重大な湧き出し」「再機械翻訳を使用」のうち一つ以上に該当する請求項を対象とした。なお、一請求項が重大な誤訳／訳抜け／湧き出しの複数に該当することはあるが、再機械翻訳が使用された場合は、重大な誤訳／訳抜け／湧き出しとの重複カウントはしていない。

2.4. 独立請求項・従属請求項別の集計結果と分析

請求項には、独立請求項と従属請求項が存在する。独立請求項は、発明の全体像を詳しく記載したもので、それ単独で発明の内容が理解できる。一方、従属請求項は、先行する独立請求項を引用し、さらなる限定を加えたものである。このため、独立請求項をふまえて理解する必要がある。

下表に、全請求項 949,882 項および事業後期分 462,367 項を独立請求項と従属請求項に分けて集計した結果を示す。

表 2.4-1. 独立請求項・従属請求項別の校正作業と重大な不備の発生率

項目		独立請求項	従属請求項
請求項数	全	142,163 項（15.0%）	807,719 項（85.0%）
	後	69,049 項（14.9%）	393,318 項（85.1%）
一請求項あたりの平均文長	全	420.4 文字	88.7 文字
	後	427.9 文字	89.4 文字
一請求項あたりの平均改行数	全	6.0 回	1.4 回
	後	6.1 回	1.4 回
校正作業が発生した請求項	全	141,520 項（99.5%）	679,108 項（84.1%）
	後	68,785 項（99.6%）	331,732 項（84.3%）
重大な不備を含む請求項	全	84,624 項（59.5%）	96,939 項（12.0%）
	後	46,342 項（67.1%）	52,611 項（13.4%）
・重大な誤訳	全	61,377 項（43.2%）	75,655 項（9.4%）
	後	34,328 項（49.7%）	41,144 項（10.5%）
・重大な訳抜け	全	44,209 項（31.1%）	24,061 項（3.0%）
	後	24,599 項（35.6%）	12,709 項（3.2%）
・重大な湧き出し	全	26,637 項（18.7%）	15,262 項（1.9%）
	後	13,100 項（19.0%）	6,477 項（1.6%）
・再機械翻訳を使用	全	3,626 項（2.6%）	1,928 項（0.2%）
	後	1,158 項（1.7%）	704 項（0.2%）

上表のとおり、独立請求項と従属請求項とで各項目の集計結果に大きな差異が生じている。また、各種不備とも独立請求項における発生率 が著しく高くなっている。これは、次の第 2.4.1 項で述べるとおり、両者の文長（文字数）の違いによるところが大きいと考えられる。

2.4.1. 原文の平均文長と重大不備の発生率との関係の考察

請求項は、原則として一文で記載される。独立請求項は、発明の全体像を詳しく説明する必要があり、これを一文で記載するため、長大かつ複雑な文構造となる傾向が強い。一方、従属請求項では、発明の全体像は「請求項 n に記載の〇〇〇であって,」といった形で独立請求項を引用して簡潔に記され、さらなる限定部分のみが具体的に記載されるため、比較的シンプルな構造の短文となる。この違いは原文の平均文長に表れており、独立請求項の全件平均文長は 420.4 文字であるのに対し、従属請求項は平均 88.7 文字で、両者には約 4.7 倍の差がある。

機械翻訳の難易度は、文長が伸びるにつれて上昇する。したがって、独立請求項が従属請求項に比べて不備が多くなるのは自然である。独立請求項の平均文長は従属請求項の約 4.7 倍であったが、事業後期における重大な不備の発生率は独立請求項で 67.1%、従属請求項では 13.4%で、その差は約 5.0 倍と、おおむね文長比と合致している。

2.4.2. 請求項内の意図的改行数に関する考察

機械翻訳は、文単位で翻訳処理を行う[23]。このため、一文で記載される請求項は、原則、項全体が一括で翻訳処理される。

一般的に中日機械翻訳では、文の認識を、主に句点（。）と改行コードにより行うと考えられる。改行を文区切りと認識することで、見出し文など文末に句点を伴わない文も適切に区切ることができる。

請求項、とりわけ長大・複雑となる傾向が強い独立請求項では、発明の構成要素やステップ単位で意図的に改行を挿入して、請求項を読み易くしていることが多い。その場合、機械翻訳処理も意図的改行単位の適度な文長で行われることとなる。この状況に鑑み、前掲の集計結果をさらに分析する。

請求項中の平均改行数[24]は、前表に示したとおり、独立請求項が 6.0、従属請求項が 1.4 であった。つまり、独立請求項の平均文長は 420.4 文字だが、実際の翻訳処理単位はこれを改行数 6.0 で割った 70.1 文字程度であると概算できる。従属請求項は 88.7 文字÷1.4≒63.4 文字であり、この試算からは、長大な独立請求項であっても、実際の機械翻訳の処理単位としては従属請求項に対して文長面での極端な不利はないことになる。

---

[23] 文が長大で一括で処理できる限度を超過する場合には、文を複数に分割して個別に翻訳することがある。

[24] 請求項中に含まれる改行コード数。項末の改行もカウントしており、最低値は 1 となる。

事実、後期分の独立請求項の重大不備の発生率 67.1%を平均改行数 6.1 で除し、実際の翻訳処理単位での発生率を約 11.0%と計算することもできる。従属請求項の重大不備の発生率 13.4%を改行数 1.4 で除すと約 9.6%となり、双方の差は大幅に縮まる。

ただし、これは「請求項 1 項あたりの不備が 1 箇所のみ」という前提での単純計算である。重大不備の発生頻度は不備の有無のみを記録しており、同一請求項中で不備が複数回発生してもカウントには反映されない。このため、実際の翻訳処理単位において独立請求項と従属請求項の重大不備の発生頻度が同程度であるかは確言できない。むしろ校正作業の印象からは、長大な独立請求項は多数の不備が併発する傾向が強く、翻訳処理単位での比較でも従属請求項より不備の発生率が高い可能性が大きい。この点は、次の第 2.4.3 項でさらに考察する。

2.4.3. 重大な不備の類型別の発生頻度の分析

重大な不備の記録は請求項単位での発生の有無を基準としており、同一請求項の複数個所で不備が発生してもカウントには反映されない。ただし、本調査では「重大な誤訳」、「重大な訳抜け」、「重大な湧き出し」、「再機械翻訳を利用」の類型別に記録を取っており、類型ごとの発生状況を独立請求項と従属請求項とで比較することができる。

まず、「重大な誤訳」の発生率（後期分）は、独立請求項で 49.7%、従属請求項では 10.5%であった。両者の比率は約 4.7：1 であり、文長比（約 4.7：1）と合致している。

一方、「重大な訳抜け」と「重大な湧き出し」は、前者が従属請求項 3.2%に対して独立請求項では 35.6%（11.1 倍）、後者が従属請求項 1.6%に対して独立請求項では 19.0%（11.9 倍）と、いずれも 10 倍以上、文長比の 2 倍強に相当する差が生じている。つまり、独立請求項は、重大な訳抜け・湧き出しの発生率が従属請求項に比べて顕著に高い。この結果から、これら二種の不備は、入力文が長くなると発生率が大幅に高まり、かつ、意図的改行で適度な文長に区切られて翻訳処理されても発生率が大幅には下がらない性質を持つと仮定される。この傾向については、第 2.5.3 項で改めて分析する。さらに第 3 章の第 3.3 項、第 3.4 項では長大な請求項の機械和訳で実際に発生した訳抜け・湧き出しの実例を示し、その具体的な状況や発生原因等を詳しく考察する。

「再機械翻訳の利用」については、独立請求項で 1.7%、従属請求項で 0.2%とともに少数ながら、両者の比率の差は 8.5 倍と開いている。本事業では、機械和訳が著しく低品質で校正困難な場合に、別の機械翻訳エンジンによる再機械翻訳結果を利用可能とした（⇒第 2.1 項）。この集計結果は、独立請求項に翻訳精度が極端に低いものが一定数存在したことを示している。その一方で、再機械翻訳結果は校正者が都度申請して取り寄せる運用と

したため、申請の手間や取得までのタイムラグから、短文の従属請求項でこの制度を利用するメリットは小さく、独立請求項に利用が偏るのは自然な結果といえる。

前項末尾で「長大な独立請求項は多数の不備が併発する傾向が強く、翻訳処理単位での比較でも従属請求項より不備の発生率が高い可能性が大きい」との仮説を述べた。これは、独立請求項と従属請求項における四種の不備類型の発生率を合算して比較することで、ある程度は検証できる。事業後期分の四種の不備の発生率の合算値は独立請求項が106.0（49.7%＋35.6%＋19.0%＋1.7%）、従属請求項が 15.5（10.5%＋3.2%＋1.6%＋0.2%）であった。これをそれぞれの意図的改行数 6.1 と 1.4 で除すと、独立請求項は17.4、従属請求項は 11.1 となる。つまり、独立請求項は従属請求項に比べ、一翻訳処理単位中で複数種の重大不備が発生する確率が約 57%高い計算となる。この結果は、上記仮説に一定の裏付けを与える。

2.4.4. まとめ

本項に示した集計結果から、現状の機械翻訳プラットフォームの機械和訳では、内容理解を阻害するレベルの重大不備は独立請求項に大きく偏って発生していることがわかった。具体的には、全独立請求項の 3 分の 2 に相当する 67.1%程度に何らかの重大な不備が生じている。さらに第 2.4.3 項での考察から、一請求項中で複数種の重大不備が併発する確率も高い。

独立請求項に重大な不備が多発した最大の要因は、長大な文長にあると考えられる。ただし、独立請求項と従属請求項では記載様式にも違いがあり、文長以外の要素も不備の多寡に関与している可能性がある。この点については、第 2.5.2 項で考察する。

一方、従属請求項は、重大な不備の発生率は全体の 13.4%程度と比較的低く、良好な翻訳精度であることが確認された。ただし、従属請求項は、引用している独立請求項の内容をふまえて理解するものであるため、従属請求項が正確に翻訳されていても、引用する独立請求項の翻訳文が理解困難であれば、事実上、従属請求項も理解はできない。このため、機械翻訳プラットフォームの実質的な翻訳精度は、独立請求項を基準に考えるのが適切であろう。

2.5. 請求項の文長に基づく集計結果と分析

前項で、独立請求項と従属請求項における重大不備の発生率の差異は、主に文長の違いに起因すると考察した。ただし、独立請求項と従属請求項では記載様式にも違いがあり、文長以外の要素も不備の発生に関与している可能性もある。本項にて、独立請求項と従属請求項をそれぞれ文長別に統計し、双方を比較してこの点を検証する。

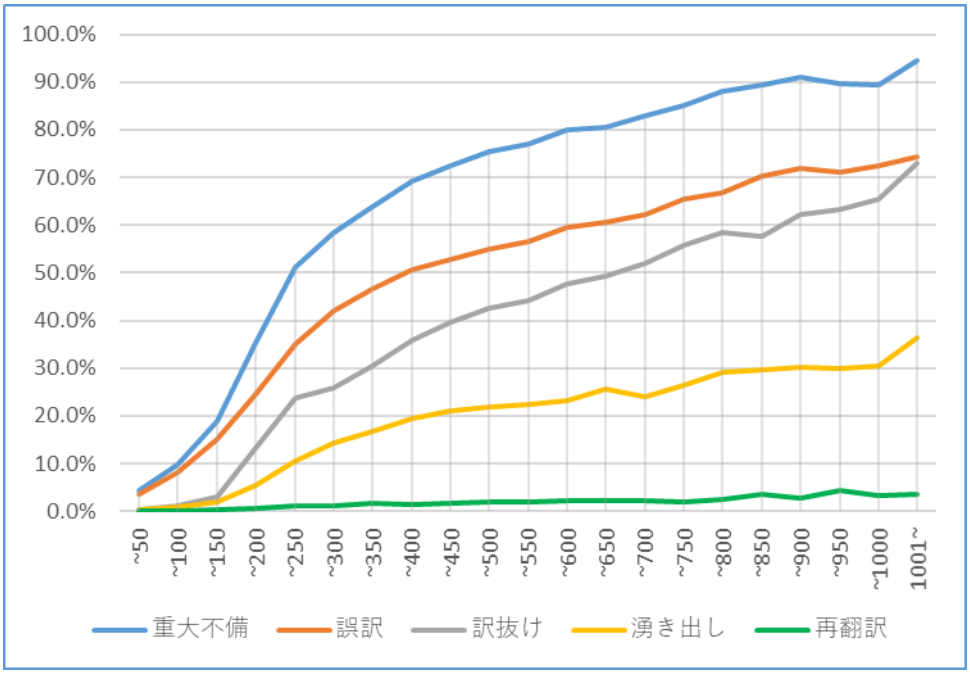
2.5.1. 全請求項の文長別統計

まずは事業後期分の全請求項を文長別に統計した結果を下表に示す。あわせて、重大な不備と、その四種の類型（重大な誤訳／重大な訳抜け／重大な湧き出し／再機械翻訳の利用）ごとの文長別発生率をグラフで比較する。

表 2.5.1-1. 請求項の文長別集計結果（事業後期分）

原文文長	請求項数（構成比）	校正あり	重大不備	（重大不備の類型別内訳）			
				誤訳	訳抜け	湧出し	再翻訳
～50	96,337	63,443	4,184	3,489	308	439	26
	(20.8%)	65.9%	4.3%	3.6%	0.3%	0.5%	0.0%
51～100	195,020	169,273	19,053	16,012	2,096	1,577	158
	(42.2%)	86.8%	9.8%	8.2%	1.1%	0.8%	0.1%
101～200	90,880	87,890	21,657	16,403	5,646	2,657	299
	(19.7%)	96.7%	23.8%	18.0%	6.2%	2.9%	0.3%
201～300	29,610	29,441	16,069	11,287	7,317	3,599	346
	(6.4%)	99.4%	54.3%	38.1%	24.7%	12.2%	1.2%
301～500	30,618	30,574	21,180	15,473	11,000	5,906	513
	(6.6%)	99.9%	69.2%	50.5%	35.9%	19.3%	1.7%
501～1,000	16,742	16,736	13,823	10,456	8,633	4,251	407
	(3.6%)	100.0%	82.6%	62.5%	51.6%	25.4%	2.4%
1,001～	3,160	3,160	2,987	2,352	2,308	1,148	113
	(0.7%)	100.0%	94.5%	74.4%	73.0%	36.3%	3.6%
全体	462,367	400,517	98,953	75,472	37,308	19,577	1,862
		86.6%	21.4%	16.3%	8.1%	4.2%	0.4%

図 2.5.1-1. 重大不備の類型ごとの文長別発生率（事業後期分）



上表・上図のとおり、重大不備の発生率は文長に相関して上昇している。各種不備とも文長が短い間は発生率は低いが、200 文字を超えた地点（つまり 201～300 文字の範囲）で「重大な誤訳」が 18.1%→38.1%（約 2.1 倍）、「重大な訳抜け」が 6.2%→24.7%（約 4.0 倍）、「重大な湧き出し」が 2.9%→12.2%（約 4.2 倍）と、いずれも直前の範囲（101～200 文字）から急増している。文字範囲をさらに細分化した上記グラフ（図 2.5.1-1）を見ると、重大な誤訳、訳抜け、湧き出しのいずれも 150 文字を超えた地点で不備の頻度が急増していることがわかる。とりわけ重大な訳抜けと重大な湧き出しは、150 文字までは非常に低頻度で横ばいに近い推移であるのに対し、150 文字を超えた時点で頻度が急激に高まっている。こうした状況から、前項で述べた「これら二種の不備は、入力文が長くなると発生率が大幅に高まる性質を有する」という考察が裏付けられるとともに、「150 文字超」という具体的な文長閾値が示唆された。

機械翻訳の精度は原則として入力文の文長と負の相関があるが、その関係性は必ずしも全文長を通じて一律とは限らない。むしろ、文分割処理が行われる文長や用語間の関係性を把握できる距離[25]、さらには記憶領域のサイズといったエンジンごとの条件により、翻訳精度が顕著に低下する閾値が存在する場合が多いと推測される。上掲のグラフから、機械翻訳プラットフォームにおいては 150 文字～250 文字の範囲がこれに該当すると見なせる。す

[25] ニューラル機械翻訳では、入力文中の各用語とその周辺の語句との関係性をベクトル化して、それに最も近似のベクトル値をもつ訳語を学習データから採用している。ただし、エンジンが用語と周辺語句との関係性を把握できる範囲（距離）には限度がある。この課題については、第 3.7.4 項で考察する。

なわち、文長がこの閾値に到ると、重大な誤訳のコンスタントな増加に加えて、訳抜け・湧き出しが急増してくる。

本調査の対象となった独立請求項の平均文長は 420.4 文字であり、上述の機械翻訳プラットフォームの閾値を超えている。したがって、長文における重大不備の改善は、現行の機械翻訳プラットフォームの最も重要な課題のひとつとなる。

2.5.2. 独立請求項／従属請求項の文長別統計

続いて、独立請求項、従属請求項それぞれの文長別の統計結果を示す。まずは独立請求項の集計結果を下表に示す。

表 2.5.2-1. 独立請求項の文長別集計結果（事業後期分）

原文文長	請求項数（独項比）	校正あり	重大不備	（重大不備の類型別内訳）			
				誤訳	訳抜け	湧出し	再翻訳
～50	191	155	33	31	1	2	0
	(0.2%)	81.2%	17.3%	16.2%	0.5%	1.0%	0.0%
51～100	1,110	1,043	290	234	44	35	1
	(0.6%)	94.0%	26.1%	21.1%	4.0%	3.2%	0.1%
101～200	8,380	8,288	3,559	2,646	1,142	634	54
	(9.2%)	98.9%	42.5%	31.6%	13.6%	7.6%	0.6%
201～300	15,691	15,650	9,023	6,507	3,861	2,179	202
	(53.0%)	99.7%	57.5%	41.5%	24.6%	13.9%	1.3%
301～500	25,224	25,200	17,755	12,970	9,260	5,119	419
	(82.4%)	99.9%	70.4%	51.4%	36.7%	20.3%	1.7%
501～1,000	15,437	15,433	12,823	9,682	8,076	4,020	373
	(92.2%)	100.0%	83.1%	62.7%	52.3%	26.0%	2.4%
1,001～	3,016	3,016	2,859	2,258	2,215	1,111	109
	(95.4%)	100.0%	94.8%	74.9%	73.4%	36.8%	3.6%
全体	69,049	68,785	46,342	34,328	24,599	13,100	1,158
	(14.9%)	99.6%	67.1%	49.7%	35.6%	19.0%	1.7%

表中「請求項数」欄には「(独項比)」として、当該文長範囲に属する全請求項に占める独立請求項の比率を示した。例えば文長「～50」の独項比 0.2%は、この文長範囲の全請求項のうち独立請求項は 0.2%であり、残る 99.8%は従属請求項であることを意味している。

独項比の集計結果から、文長 200 文字以下の範囲では独立請求項はごく少数であり、反面、文長が 300 文字を超えるものは大半が独立請求項であることがわかる。その中間に位置する「201～300 文字」の範囲が、独立請求項が 53.0%、従属請求項は 47.0%と、両者が最も均等に分布するゾーンとなっている。

続いて、従属請求項のみの文長別統計を下表に示す。

表 2.5.2-2. 従属請求項の文長別集計結果（事業後期分）

原文文長	請求項数（従項比）	校正あり	重大不備	（重大不備の類型別内訳）			
				誤訳	訳抜け	湧出し	再翻訳
～50	96,146	63,288	4,151	3,458	307	437	26
	(99.8%)	65.8%	4.3%	3.6%	0.3%	0.5%	0.0%
51～100	193,910	168,230	18,763	15,778	2,052	1,542	157
	(99.4%)	86.8%	9.7%	8.1%	1.1%	0.8%	0.1%
101～200	82,500	79,602	18,098	13,757	4,504	2,023	245
	(90.8%)	96.5%	21.9%	16.7%	5.5%	2.5%	0.3%
201～300	13,919	13,791	7,046	4,780	3,456	1,420	144
	(47.0%)	99.1%	50.6%	34.3%	24.8%	10.2%	1.0%
301～500	5,394	5,374	3,425	2,503	1,740	787	94
	(17.6%)	99.6%	63.5%	46.4%	32.3%	14.6%	1.7%
501～1,000	1,305	1,303	1,000	774	557	231	34
	(7.8%)	99.8%	76.6%	59.3%	42.7%	17.7%	2.6%
1,001～	144	144	128	94	93	37	4
	(4.6%)	100.0%	88.9%	65.3%	64.6%	25.7%	2.8%
全体	393,318	331,732	52,611	41,144	12,709	6,477	704
	(85.1%)	84.3%	13.4%	10.5%	3.2%	1.6%	0.2%

あわせて、独立請求項、従属請求項それぞれの、4 種の重大不備の類型ごとの文長別発生率をグラフで表示する。

図 2.5.2-1. 独立請求項の重大不備の類型ごとの文長別発生率（事業後期分）

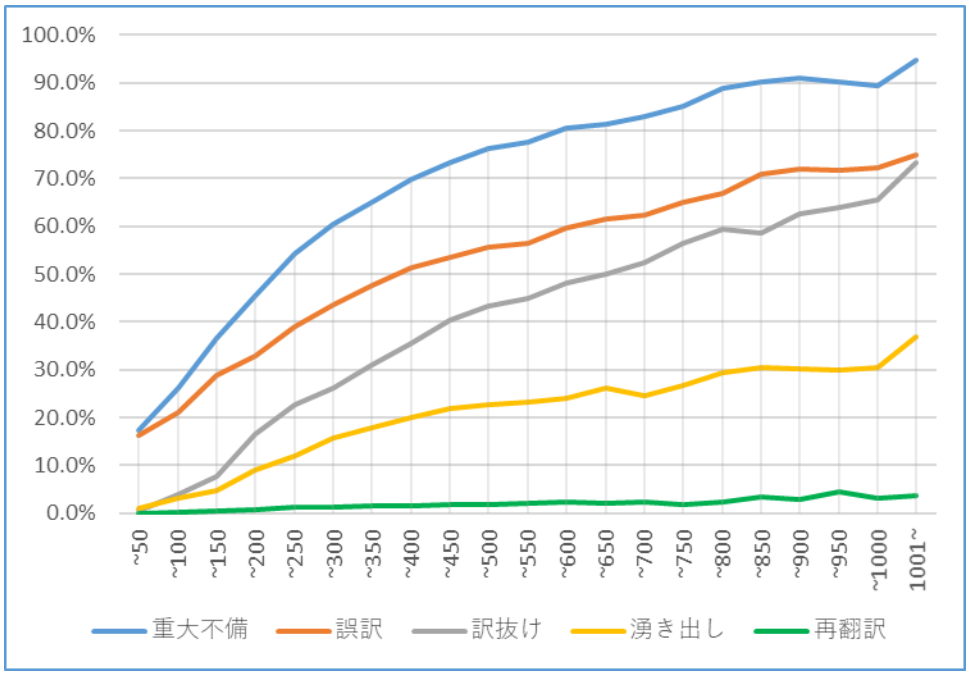
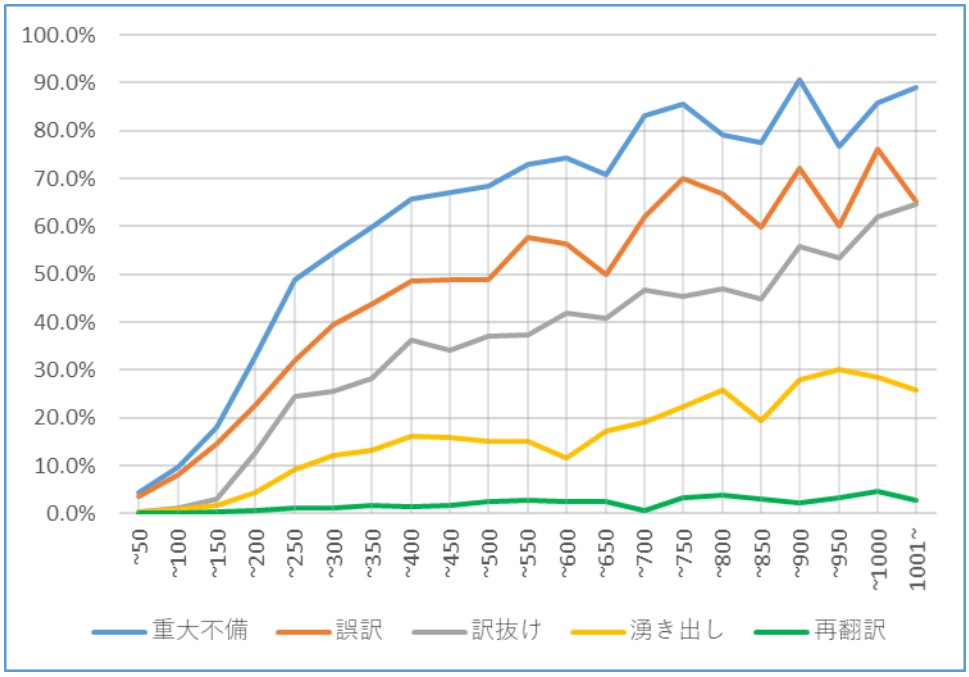


図 2.5.2-2. 従属請求項の重大不備の類型ごとの文長別発生率（事業後期分）



独立請求項、従属請求項それぞれの統計結果を対比すると、両者とも文長が伸びるにつれて重大な不備の発生率が増加しており、各文長範囲における発生率も、大枠としては同程度といえる[26]。つまり、第 2.4 項で考察したとおり、独立請求項と従属請求項の不備発生頻度の大きな差は主に文長の構成比の違いに起因していることがわかる。

ただしその一方で、全ての文長範囲において独立請求項のほうが不備の発生率がやや高く、その差は文長が短いほど顕著となることも判明した。下表に、独立請求項と従属請求項の各文長範囲での重大不備の発生率と双方の頻度差を一覧にまとめた。

表 2.5.2-3. 独立請求項と従属請求項の不備発生率の文長別比較（事業後期分）

文長	～50	51～100	101～200	201～300	301～500	501～1,000	1,001～
独立項	33	290	3,559	9,023	17,755	12,823	2,859
	17.3%	26.1%	42.5%	57.5%	70.4%	83.1%	94.8%
従属項	4,151	18,763	18,098	7,046	3,425	1,000	128
	4.3%	9.7%	21.9%	50.6%	63.5%	76.6%	88.9%
独 vs 従	+13.0%	+16.4%	+20.6%	+6.9%	+6.9%	+6.5%	+5.9%
頻度差	4.02 倍	2.69 倍	1.94 倍	1.14 倍	1.11 倍	1.08 倍	1.07 倍

上表のとおり、文長が短い間は、独立請求項と従属請求項の重大不備の発生率の差は顕著であり、例えば「～50 文字」では、発生率で+13.0 ポイント、頻度差では約 4.02 倍、独立請求項のほうが重大不備が多く発生している。この頻度差は文長が伸びるにつれ縮まっていき、「201～300 文字」以降は+6～7 ポイント前後、約 1.1 倍前後で安定するものの、全ての文長範囲において独立請求項のほうが発生率が高くなっている。

この状況から、従属請求項には、独立請求項に比べて不備の発生が抑制される何らかの性質を有し、その影響は、文長が短いほど顕著となることがうかがえる。その性質とは、具体的には、従属請求項の文型がパターン化されていることであると考えられる。

従属請求項では、たとえば「根据权利要求 XX 所述的○○○，其特征在于，包括：…（請求項 XX に記載の○○○であって,さらに以下を含むことを特徴とする：…)」といった定番表現が案件単位で統一的に使用される。過去事例を学習して翻訳文を生成するニューラル

[26] 従属請求項のグラフ（図 2.5.2-2）では文長が 500 文字を超えるあたりから不備発生率の変動が激しくなっているが、これは従属請求項に長大なものが少なく、文長範囲を細分化したことで数値が粗くなったためと考えられる。

機械翻訳方式は定番表現を得意としており、このため従属請求項は、確固とした定型パターンが存在しない独立請求項に比べて、同等の文長での翻訳難易度が低くなると考えられる[27]。特に短文の請求項では、全体に占める定型パターン部分の割合が高まるため、不備の発生率が大幅に抑制されたものと考えられる。

とはいえ、独立請求項、従属請求項とも文が長大になるにつれて重大不備の発生率は急激に上昇していき、かつ 200 文字を超えた時点からは、双方の頻度差は 1.1 倍程度とほぼ同水準に接近する。この結果から、基本的には請求項の文長という物理的条件が、独立／従属の種別を問わず、機械翻訳プラットフォームの翻訳精度に非常に強い影響を与えていると結論される。

2.5.3. 意図的改行・区切り記号の有無

機械翻訳では、翻訳処理の基本単位は文である。入力された文章は句点「。」で文単位に区切られ、個別に翻訳処理が行われる。また、文末の改行も区切りと見なされる。さらに、機械翻訳では、文が長大で分割処理が必要となるような場合、セミコロン「；」も区切りとして扱われることが多い。

前項で示した文長別集計結果は、請求項の文長に基づく。請求項は原則として一文で記載され、末尾以外に句点が存在するケースは稀だが、長大な請求項には意図的改行やセミコロンが挿入されることが多い。よって、同じ文長の請求項であっても、意図的改行やセミコロンで適度に区切られた請求項と、こうした区切りが存在しない請求項とでは、実質的な翻訳難易度に差がある可能性が高い。

これを検証するため、独立請求項のボリュームゾーンである文長範囲 301～500 文字に属する 25,224 項をサンプルとして、請求項中の改行の回数別に不備の発生頻度を集計した。次表・次図にその結果を示す。

---

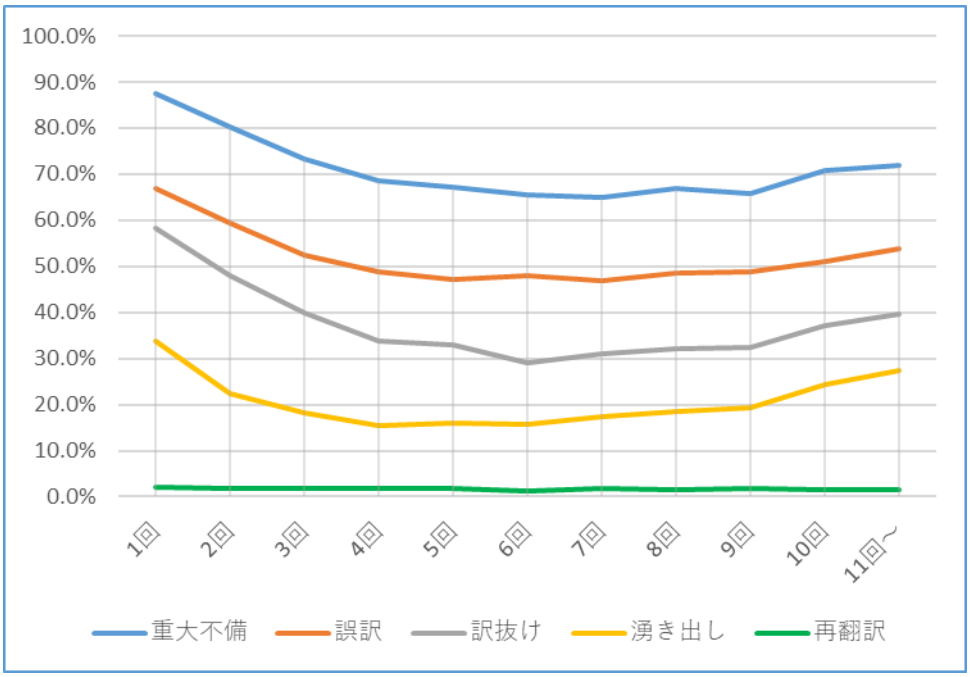
[27] 前ページのグラフを見ると、文長が短い間は特に「重大な誤訳」の発生頻度において独立請求項と従属請求項との間に傾向の差異が見られる。これは、特段の定型パターンが存在しない独立請求項は比較的文長（≒語句数）に比例して誤訳が発生するのに対し、従属請求項は文が短いほど(翻訳精度の高い)定型パターンの占める比率が高まり、（誤訳の発生する可能性のある）その他の語句の数が少量となるためと考えられる。

表 2.5.3-1. 301～500 文字の独立請求項の改行回数別の不備発生率（事業後期分）

改行数[28]	請求項数	校正	重大	（重大不備の類型別内訳）			
		あり	不備	誤訳	訳抜け	湧出し	再翻訳
1 回	3,020	3,017	2,645	2,020	1,756	1,017	60
		99.9%	87.6%	66.9%	58.1%	33.7%	2.0%
2 回	797	796	639	473	382	178	14
		99.9%	80.2%	59.3%	47.9%	22.3%	1.8%
3 回	1,436	1,431	1,051	752	574	262	26
		99.7%	73.2%	52.4%	40.0%	18.2%	1.8%
4 回	2,494	2,491	1,709	1,217	843	388	43
		99.9%	68.5%	48.8%	33.8%	15.6%	1.7%
5 回	3,440	3,438	2,314	1,626	1,131	552	61
		99.9%	67.3%	47.3%	32.9%	16.0%	1.8%
6 回	3,484	3,477	2,284	1,676	1,011	552	44
		99.8%	65.6%	48.1%	29.0%	15.8%	1.3%
7 回	3,120	3,119	2,024	1,461	969	545	56
		100.0%	64.9%	46.8%	31.1%	17.5%	1.8%
8 回	2,579	2,578	1,725	1,252	832	474	37
		100.0%	66.9%	48.5%	32.3%	18.4%	1.4%
9 回	1,810	1,810	1,189	885	584	352	31
		100.0%	65.7%	48.9%	32.3%	19.4%	1.7%
10 回	1,190	1,190	842	609	442	291	19
		100.0%	70.8%	51.2%	37.1%	24.5%	1.6%
11 回～	1,854	1,853	1,333	999	736	508	28
		99.9%	71.9%	53.9%	39.7%	27.4%	1.5%
合計	25,224	25,200	17,755	12,970	9,260	5,119	419
		99.9%	70.4%	51.4%	36.7%	20.3%	1.7%

[28] 請求項中に含まれる改行コード数。項末の改行もカウントしており、最低値は 1 となる。

図 2.5.3-1. 301～500 文字の独立請求項の改行回数別の重大不備発生率（事業後期分）



上表のとおり、重大不備の発生率は改行数 1 回（＝請求項中の意図的改行はゼロ）の時が最多（87.6%）で、意図的改行数の増加とともに減少している。ただし、減少の度合いは改行数が増えるにつれて小さくなり、8 回を超えると発生率はむしろ微増傾向に転じる。この傾向は、上図に示されるとおり、不備の類型別に見てもほぼ共通する。

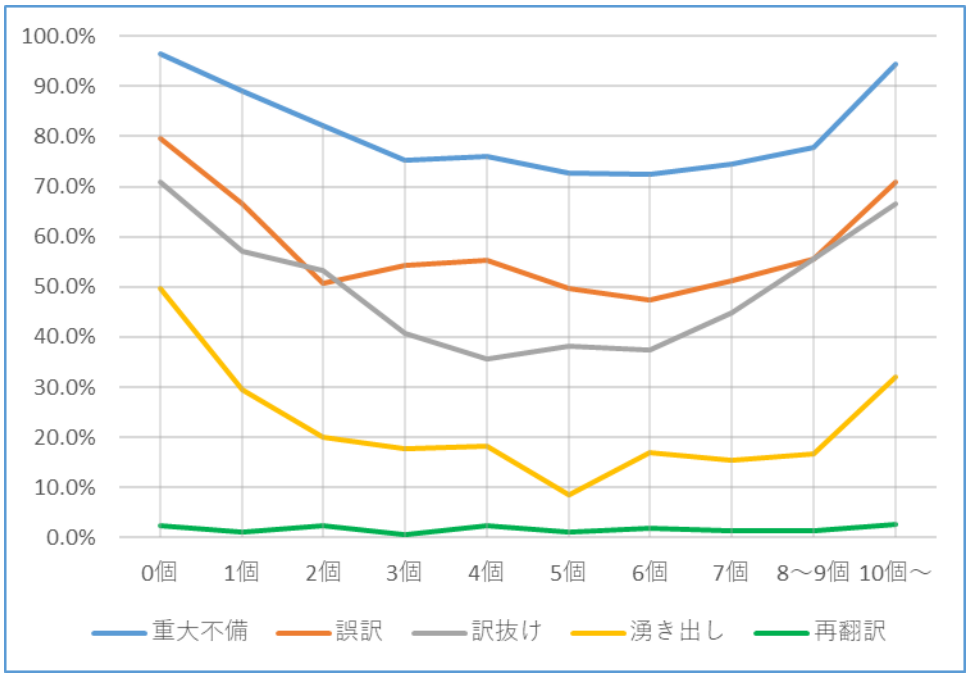
この結果から、文長 301～500 文字の範囲内の平均的なサイズの独立請求項においては、改行数が 1 回、つまり中間に意図的改行が挿入されていない請求項に比べ、適度に意図的改行が挿入された請求項のほうが、不備の発生が抑えられていると結論できる。この結果は予測したとおりではあるが、ただし、意図的改行によって翻訳処理単位の平均文長が数分の一となることに比べると、重大不備の減少度はさほど急激ではない。

これは、意図的改行が用いられない請求項であっても、エンジンがセミコロンを区切りと見なして自動的に文を分割しているためである可能性がある。これを調査するため、表中の改行数 1 回の（つまり意図的改行を含まない）請求項 3,020 項について、文中に含まれるセミコロン数ごとの集計を行った。以下、結果を示す。

表 2.5-3-2. 文長 301～500×意図的改行ゼロの独立請求項のセミコロン数別不備発生率（事業後期分）

セミコロン数	請求項数	校正あり	重大不備	（重大不備の類型別内訳）			
				誤訳	訳抜け	湧出し	再翻訳
0 個	1,377	1,376	1,328	1,097	975	684	34
		99.9%	96.4%	79.7%	70.8%	49.7%	2.5%
1 個	360	360	321	240	206	106	4
		100.0%	89.2%	66.7%	57.2%	29.4%	1.1%
2 個	304	304	250	154	162	61	7
		100.0%	82.2%	50.7%	53.3%	20.1%	2.3%
3 個	272	270	205	148	111	48	2
		99.3%	75.4%	54.4%	40.8%	17.6%	0.7%
4 個	208	208	158	115	74	38	5
		100.0%	76.0%	55.3%	35.6%	18.3%	2.4%
5 個	165	165	120	82	63	14	2
		100.0%	72.7%	49.7%	38.2%	8.5%	1.2%
6 個	112	112	81	53	42	19	2
		100.0%	72.3%	47.3%	37.5%	17.0%	1.8%
7 個	78	78	58	40	35	12	1
		100.0%	74.4%	51.3%	44.9%	15.4%	1.3%
8～9 個	72	72	56	40	40	12	1
		100.0%	77.8%	55.6%	55.6%	16.7%	1.4%
10 個～	72	72	68	51	48	23	2
		100.0%	94.4%	70.8%	66.7%	31.9%	2.8%
合計	3,020	3,017	2,645	2,020	1,756	1,017	60
		99.9%	87.6%	66.9%	58.1%	33.7%	2.0%

図 2.5.3-2. 文長 301～500×意図的改行ゼロの独立請求項のセミコロン数別不備発生率（事業後期分）



上表のとおり、意図的改行を含まない文長 301～500 文字の独立請求項 3,009 項のうち、セミコロンも全く含まないものは半数弱の 1,377 項（45.6%）存在し、その重大不備の発生率は 96.4%であった。これに対し、セミコロンを 1 つ以上含む 1,643 項の重大不備の発生率は 80.2%であった。類型別のグラフを見ても、セミコロンを含まないものが不備の発生率が最も高く、セミコロン数が増えるにつれて不備発生率は減少していくが、7 個前後を境に再び増加傾向に転じるという、意図的改行数とほぼ近似の傾向が表れた。この結果から、機械翻訳プラットフォームでは、長大な請求項を必要に応じてセミコロンで区切って翻訳しており、それが翻訳精度に一定の効果をもたらしていることが示された。

とはいえ、文長 301～500 文字の独立請求項における重大不備の発生率は、意図的改行やセミコロンが適度に使用されていても依然として 60%を超えている。これは、現行の機械翻訳プラットフォームにおいて、長文を改行やセミコロンで区切るという措置のみでは解決されない重大不備が多数存在することを意味している。第 2.4.3 項にて、重大な訳抜け／湧き出しは、「（入力文が長くなると発生率が大幅に高まり、かつ）意図的改行で適度な文長に区切られて翻訳処理されても発生率が大幅には下がらない性質を持つ」と仮定したが、本項の結果はこれと整合する。これら長大な請求項における典型的な不備の具体例と、その改善策については、第 3 章の第 3.3 項、第 3.7.4 項でさらに考察する。

2.6. IPC セクション別の集計結果と分析

中国特許公報には、発明の属する技術分野に応じて IPC（国際特許分類）が付与されている。本調査では、機械和訳における不備の発生状況に技術分野ごとの顕著な違いがあるかを把握すべく、全請求項を筆頭 IPC のセクション（A～H）別に集計・比較した。

比較を容易にするため、校正対象の公報 10 万件は、各セクションから 12,500 件ずつ、各セクションの対象案件の「特許請求の範囲」の平均文字数が 1,300 文字となるように選定している（⇒第 1.3 項）。ただし、そこに含まれる請求項の数や文長、独立／従属請求項の構成比等については特に調整していないため、各セクションの請求項数や平均文長には差異が生じた。こうした差異は、（平均文字数 1,300 文字という一定の条件下ではあるが）実際の中国特許の状況が反映されたものと見なせる。

本項で、重大な不備の発生頻度を筆頭 IPC のセクション別に集計した結果を分析する。なお、第 2.2 項で述べたとおり、重大な不備の判定には当初個人差が生じていた。こうした個人差のセクション間への影響を極力排除するため、本項では、個人差が解消した事業後期分の請求項、全 462,367 項の集計結果のみを分析対象とする。

2.6.1. 全請求項の集計結果

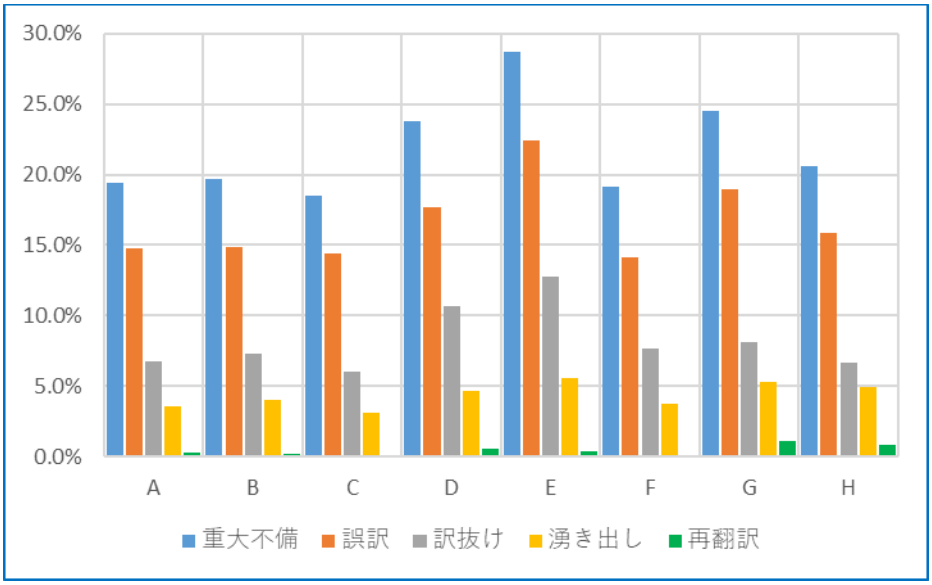
まずは事業後期分の全請求項 462,367 項をセクション別に集計した結果を示す。

表 2.6.1-1. IPC セクション別 全請求項の校正と重大不備の発生率（事業後期分）

IPC	請求項数	平均文長	校正あり	重大不備	（重大不備の類型別内訳）			
					誤訳	訳抜け	湧出し	再翻訳
A	63,130	121.4	52,688	12,272	9,337	4,263	2,243	213
			83.5%	19.4%	14.8%	6.8%	3.6%	0.3%
B	64,986	130.6	56,673	12,825	9,661	4,745	2,580	148
			87.2%	19.7%	14.9%	7.3%	4.0%	0.2%
C	73,362	124.0	59,511	13,565	10,580	4,375	2,286	110
			81.1%	18.5%	14.4%	6.0%	3.1%	0.1%
D	47,910	162.0	41,604	11,422	8,472	5,103	2,240	274
			86.8%	23.8%	17.7%	10.7%	4.7%	0.6%
E	52,545	177.7	49,034	15,104	11,750	6,733	2,936	197
			93.3%	28.7%	22.4%	12.8%	5.6%	0.4%
F	66,749	133.5	59,379	12,746	9,403	5,170	2,507	38
			89.0%	19.1%	14.1%	7.7%	3.8%	0.1%

G	44,262	151.7	39,070	10,831	8,422	3,592	2,359	477
			88.3%	24.5%	19.0%	8.1%	5.3%	1.1%
H	49,423	136.4	42,558	10,188	7,847	3,327	2,426	405
			86.1%	20.6%	15.9%	6.7%	4.9%	0.8%
全体	462,367	139.9	400,517	98,953	75,472	37,308	19,577	1,862
			86.6%	21.4%	16.3%	8.1%	4.2%	0.4%

図 2.6.1-1. IPC セクション別 全請求項の重大不備の類型別発生率（事業後期分）



各セクションの状況を比較すると、重大な不備の発生率は E セクションが 28.7%と最も高く、次いで G セクションが 24.5%、D セクションが 23.8%と続く。残る 5 セクションはいずれも平均値 21.4%を下回り、第 4 位が H セクションの 20.6%、第 8 位が C セクションの 18.5%と狭い範囲に密集した。

不備の類型ごとの集計結果も、おおむね各類型とも E、G、D セクションの順に高く、他の 5 セクションは平均値を下回るという「重大な不備」全体の傾向とほぼ合致した。わずかな例外は、「重大な訳抜け」で D と G の順位が入れ替わったことと、「重大な湧き出し」で H セクションが 4.9%で第 3 位となり、第 4 位の D セクション（4.7%）や全セクション平均値（4.2%）を上回ったこと程度であった。

E、G、D セクションで重大な不備が多発した最大の理由として、請求項の平均文長が他に比して長いことが挙げられる。具体的には、全セクションの平均文長 139.9 文字に対し、E セクションは 177.7 文字、G セクションは 151.7 文字、D セクションは 162.0 文字であった。一方、他の 5 セクションはいずれも全件平均文長を下回っている。

文長と重大不備の発生率との密接な相関性は前項で分析したとおりであるが、このようにセクション別の統計でも、両者の間には高い相関関係が見られた。8 セクションを平均文長の長い順に並べると E、D、G、H、F、B、C、A となるが、それぞれの重大不備の発生率の順位は 1 、3、2、 4 、7、5、8、6 位であり、おおむね文長に準じている。

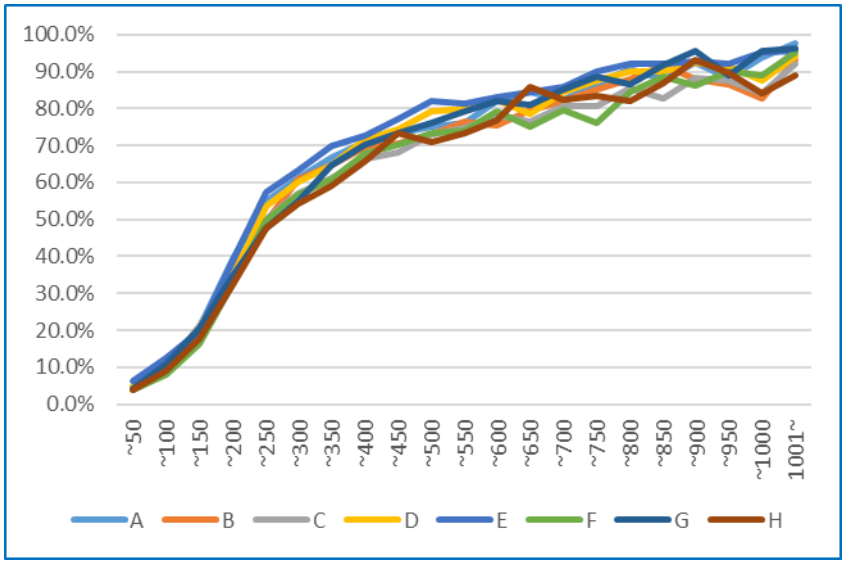
各セクションの重大不備の発生率の違いが主として文長に起因するものであるかは、各セクションにおける重大不備発生率を文長別に区分して比較することで検証できる。次表に、IPC セクション×文長別の重大不備の発生率を示す。

表 2.6.1-2. IPC セクション×文長別 重大不備発生率（事業後期分）

原文文長	全件	A	B	C	D	E	F	G	H
～50	4,185	802	609	896	314	400	460	317	386
	4.3%	4.5%	4.3%	4.1%	4.8%	6.4%	3.8%	4.2%	3.8%
51～100	19,055	2,666	2,611	2,806	1,932	2,829	2,444	1,934	1,831
	9.8%	10.2%	9.1%	9.6%	9.5%	12.7%	8.1%	10.9%	9.0%
101～200	21,663	2,735	2,854	2,974	2,562	3,128	2,851	2,377	2,176
	23.8%	26.2%	23.7%	25.8%	22.1%	25.3%	20.8%	24.9%	22.5%
201～300	16,071	1,872	2,031	2,163	1,867	2,230	2,103	1,803	2,000
	54.3%	57.5%	54.4%	53.0%	56.2%	59.9%	52.6%	50.9%	50.5%
301～500	21,180	2,435	2,808	2,494	2,234	2,999	3,034	2,659	2,517
	69.2%	70.8%	68.6%	66.2%	71.4%	74.6%	67.0%	69.9%	65.7%
501～1,000	13,823	1,475	1,629	1,803	1,936	2,673	1,623	1,528	1,156
	82.6%	83.4%	80.8%	79.8%	84.3%	86.9%	78.6%	83.9%	80.2%
1,001～	2,987	287	283	429	577	845	231	213	122
	94.5%	97.6%	93.1%	92.1%	94.3%	95.7%	95.1%	96.4%	89.1%

※黄色セルが当該文長範囲内での最多頻度、橙セルが第 2 位、緑セルが第 3 位。

図 2.6.1-2. IPC セクション×文長別 重大不備発生率（事業後期分）



上表及び上図のとおり、各セクションの重大不備発生率は各文長範囲でおおむね一定の範囲内に収まっており、全セクション同一の推移を示した。セクション間の順位も、E セクションが多くの文長範囲で最高頻度となるなど若干の序列は見られるものの、各文字範囲で少なからず変動しており、他から突出して高い／低い頻度のセクションは見られない。むしろ上掲のグラフから、文長と不備発生率の相関関係は基本的には全セクション共通といえる。つまり、セクション間の重大不備の発生率の差異が主に文長差に起因していることを肯定する結果である。

なお、セクションごとの不備の発生率に違いが生じる別の要因として、学習データ量に差がある可能性も考えられる。機械翻訳プラットフォームの学習データの詳細は非公開であるが、主に中日対応特許由来のコーパスが使用されていると推測される。当財団は、2021 年発行分までの中日対応特許のほぼ全件に由来する約 2.3 億文対の対訳コーパスを保有しており、そのセクション別の文対数が、現時点で入手可能な各セクションの中日対応特許コーパスの総数と見なせる。以下に示す。

表 2.6.1-3. 当財団中日対応特許コーパスの IPC セクション別文対数（単位：万件）

ｾｸｼｮﾝ	A	B	C	D	E	F	G	H
文対数	3,680	2,812	4,616	199	174	1,394	4,440	6,037
構成比	15.8%	12.0%	19.8%	0.9%	0.7%	6.0%	19.0%	25.9%

上表に示したとおり、E セクションと D セクションは、他に比べて中日対応特許コーパスの件数が非常に少ない。つまり、両セクションは、中日対応特許由来のデータの流通量

自体が他に比して少ないといえる。このため、学習データの充実度が相対的に他のセクションより劣り、このため重大な不備の発生率が高くなっている可能性がある。

ただし、E、D セクションと同様に不備の発生率が高かった G セクションのコーパス件数は 4,440 万件（全セクション中第 2 位）と多いなど、この仮説には矛盾点もあり、文長ほどの明瞭な相関性は見られない。機械翻訳プラットフォームの学習データが上掲の比率に準じた構成であるかも不明であり、参考程度に取り扱うべき情報であろう。

続いて、実際の中国特許のセクション別の構成比についても調査した。具体的には、2021 年の中国公開公報 1,720,484 件の筆頭 IPC 別構成比を調査した。下表に示す。

表 2.6.1-4. 2021 年中国公開公報の IPC セクション別文献数（単位：万件）

セクション	A	B	C	D	E	F	G	H
文献数	20.0	35.8	20.0	2.4	8.3	11.4	47.7	26.5
構成比	11.6%	20.8%	11.6%	1.4%	4.8%	6.6%	27.7%	15.4%

中国公開公報のセクション構成比も、D セクションが最も少なく、次いで E セクション、F セクションの順と、おおむね対応特許由来コーパスの文対数と合致している。中国での出願件数の多寡が、中日対応特許の多寡にほぼ直結していることがわかる。

機械翻訳プラットフォームにおいて、D、E セクションの翻訳精度を他のセクションと同程度に引き上げる手段としては、学習データを増強することが考えられる。ただし、両セクションは中日対応特許自体が他のセクションに比べて極めて少量であり、現在流通している以上の学習データを得るには多大なコストが必要となる。かつ、機械翻訳プラットフォームのように全セクションの学習データを一つのエンジンに学習させている場合、特定のセクションのみを重点的に学習させると、そのセクションの翻訳傾向が他のセクションに必要以上に反映され、悪影響を及ぼす懸念も皆無ではない。したがって、この方策が全体的な翻訳精度向上となるかは確言できない。

上掲した中国公開公報のセクション別構成比は、事実上、機械翻訳プラットフォームで機械和訳の対象となる中国公報の分野構成比であるといえる。その意味では、この構成比に準じた学習データを用いるほうが、分野間のバランスの観点からは理に適っているとも考えられる。

本項に示した結果からは、セクション間の翻訳精度の差は、主に文長差に起因している可能性が高い。したがって、前項までの結論と同様、長大な請求項の翻訳精度を向上させ

ることが、セクション間の翻訳品質の差を解消することにもつながると考える。

2.6.2. 独立請求項／従属請求項別の集計結果

本項では、各 IPC セクションに属する請求項をさらに独立請求項／従属請求項別に集計した結果を分析する。まずは集計結果を下二表に示す。なお、各表「請求項数」欄の 2 段目には、当該セクションにおける独立請求項／従属請求項の構成比を示した。

表 2.6.2-1. IPC セクション別 独立請求項の校正と重大不備の発生率（事業後期分）

IPC	請求項数 （独項比）	平均 文長	校正 あり	重大 不備	（重大不備の類型別内訳）			
					誤訳	訳抜け	湧出し	再翻訳
A	8,397 (13.3%)	392.4	8,353	5,600	4,071	2,802	1,411	150
			99.5%	66.7%	48.5%	33.4%	16.8%	1.8%
B	9,018 (13.9%)	414.5	8,995	5,982	4,381	3,153	1,727	77
			99.7%	66.3%	48.6%	35.0%	19.2%	0.9%
C	9,597 (13.1%)	393.5	9,481	5,951	4,523	2,756	1,333	77
			98.8%	62.0%	47.1%	28.7%	13.9%	0.8%
D	6,697 (14.0%)	521.6	6,664	4,884	3,532	3,031	1,358	163
			99.5%	72.9%	52.7%	45.3%	20.3%	2.4%
E	8,250 (15.7%)	558.7	8,245	6,509	4,997	4,246	2,064	41
			99.9%	78.9%	60.6%	51.5%	25.0%	0.5%
F	9,432 (14.1%)	412.0	9,425	6,144	4,360	3,456	1,768	30
			99.9%	65.1%	46.2%	36.6%	18.7%	0.3%
G	8,496 (19.2%)	403.0	8,478	5,698	4,295	2,659	1,713	330
			99.8%	67.1%	50.6%	31.3%	20.2%	3.9%
H	9,162 (18.5%)	362.6	9,144	5,574	4,169	2,496	1,726	290
			99.8%	60.8%	45.5%	27.2%	18.8%	3.2%
全体	69,049 (14.9%)	427.9	68,785	46,342	34,328	24,599	13,100	1,158
			99.6%	67.1%	49.7%	35.6%	19.0%	1.7%

表 2.6.2-2. IPC セクション別 従属請求項の校正と重大不備の発生率（事業後期分）

IPC	請求項数 （従項比）	平均 文長	校正 あり	重大 不備	（重大不備の類型別内訳）			
					誤訳	訳抜け	湧出し	再翻訳
A	54,733 (86.7%)	79.8	44,335	6,672	5,266	1,461	832	63
			81.0%	12.2%	9.6%	2.7%	1.5%	0.1%
B	55,968 (86.1%)	84.8	47,678	6,843	5,280	1,592	853	71
			85.2%	12.2%	9.4%	2.8%	1.5%	0.1%
C	63,765 (86.9%)	83.5	50,030	7,614	6,057	1,619	953	33
			78.5%	11.9%	9.5%	2.5%	1.5%	0.1%
D	41,213 (86.0%)	103.5	34,940	6,538	4,940	2,072	882	111
			84.8%	15.9%	12.0%	5.0%	2.1%	0.3%
E	44,295 (84.3%)	106.7	40,789	8,595	6,753	2,487	872	156
			92.1%	19.4%	15.2%	5.6%	2.0%	0.4%
F	57,317 (85.9%)	87.7	49,954	6,602	5,043	1,714	739	8
			87.2%	11.5%	8.8%	3.0%	1.3%	0.0%
G	35,766 (80.8%)	92.0	30,592	5,133	4,127	933	646	147
			85.5%	14.4%	11.5%	2.6%	1.8%	0.4%
H	40,261 (81.5%)	84.9	33,414	4,614	3,678	831	700	115
			83.0%	11.5%	9.1%	2.1%	1.7%	0.3%
全体	393,318 (85.1%)	89.4	331,732	52,622	41,152	12,712	6,477	704
			84.3%	13.4%	10.5%	3.2%	1.6%	0.2%

図 2.6.2-1. IPC セクション別 独立請求項の重大不備の類型別発生率（事業後期分）

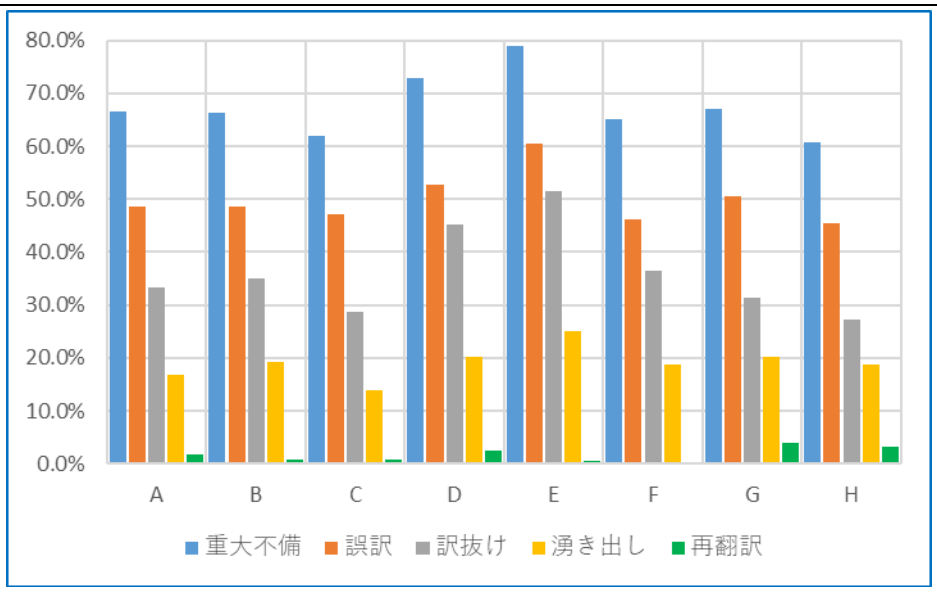
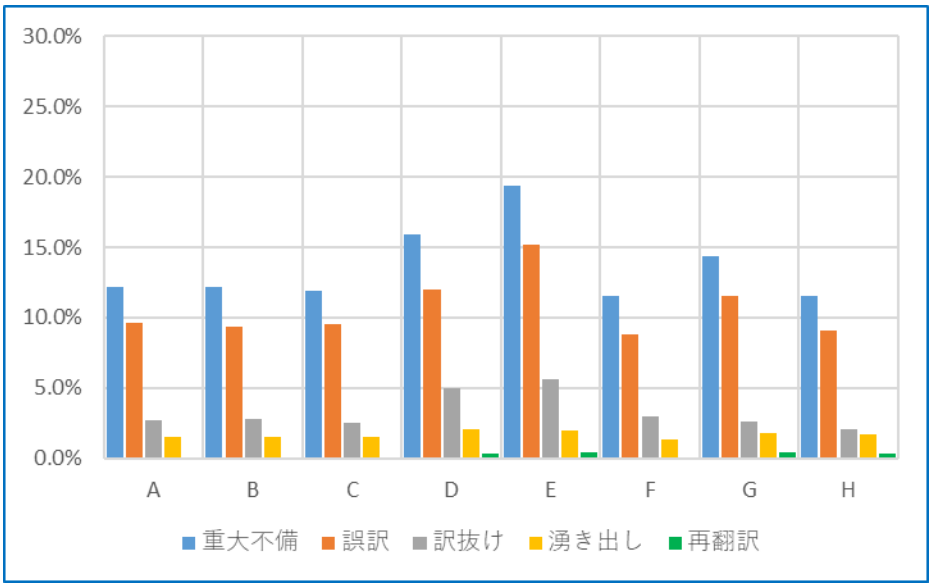


図 2.6.2-2. IPC セクション別 従属請求項の重大不備の類型別発生率（事業後期分）



独立請求項では、重大不備の発生率は E セクション（78.9%）と D セクション（72.9%）が特に高く、全請求項では重大不備発生率が第 2 位であった G セクションは、独立請求項に限れば発生率は平均値と同率の 67.1%であった。残る 5 セクションは全て平均値を下回っており、発生率が最も低かったのは H セクション（60.8%）であった。

E セクションと D セクションは、独立請求項のみの比較でも、全請求項での比較と同様、他のセクションに比べて平均文長が突出して長い。具体的には E が 558.7 文字、D が

521.6 文字であった。他の 6 セクションはいずれも平均値（427.9 文字）以下であり、この文長の違いが E、D セクションの不備の発生率を高めている大きな要因であることは確実である。

一方、独立請求項の構成比に関しては、D セクションは 14.0%、E セクションは 15.7%と、いずれも全体平均（14.9%）から±1%程度の差異にとどまっている。むしろ G セクション（19.2%）と H セクション（18.5%）で独立請求項の構成比が高い。独立請求項の翻訳難易度が高いことは前々項で述べたとおりであり、両セクションが全請求項の重大不備率で上位（第 2 位、第 4 位）となった一因と考えられる。

従属請求項の集計結果においても、平均文長と重大不備の発生率は明瞭な相関関係にあった。従属請求項のみの比較で文長が全体平均 89.4 文字よりも長いセクションは E セクション（106.7 文字)、D セクション（103.5 文字)、G セクション（92.0 文字）の 3 セクションであったが、それぞれの重大不備発生率も E が 19.4%（第 1 位)、D が 15.9%（第 2 位)、G が 14.4%（第 3 位）と上位を占めた。

全請求項に占める従属請求項の構成比では、これら 4 セクションのうち G（80.8%)、H（81.5%)、E（84.3%）の 3 セクションは全体平均（85.1%）を下回っている（D は 86.0%)。翻訳難易度の低い従属請求項の構成比が小さければ、他セクションとの比較で不利に働くことは言うまでもない。

本調査により、独立請求項が他に比して極端に長大な傾向となるセクション（E、D）と、独立請求項の構成比が他に比して高いセクション（G、H）で、重大不備の発生率が高くなっていることが明らかになった。前項までで考察したとおり、機械翻訳プラットフォームの翻訳品質が、長大な独立請求項の翻訳精度に大きく左右されている状況がさらに明確に示された。

2.7. 出願人の国コード別の集計結果と分析

中国特許公報の書誌情報に含まれる「出願人の国コード」により、中国（CN）からの出願と、中国以外の諸外国からの出願とで機械和訳の翻訳精度に違いがあるかを調査した。さらに、中国以外の諸外国のうち日本（JP）と米国（US）について個別に集計した。下表に結果を示す。

表 2.7-1. 出願人国コード別 校正と重大不備の発生率（事業後期分）

国	請求項数	平均文長	校正あり	重大不備	（重大不備の類型別内訳）			
					誤訳	訳抜け	湧出し	再翻訳
CN	205,785	177.6	184,224	56,767	43,705	23,989	11,546	1,260
			89.5%	27.6%	21.2%	11.7%	5.6%	0.6%
CN以外	256,582	109.8	216,293	42,186	31,767	13,319	8,031	602
			84.3%	16.4%	12.4%	5.2%	3.1%	0.2%
(JP)	38,529	133.4	34,271	6,991	5,107	1,951	1,445	102
			88.9%	18.1%	13.3%	5.1%	3.8%	0.3%
(US)	85,457	100.8	71,112	13,975	10,584	4,287	2,426	213
			83.2%	16.4%	12.4%	5.0%	2.8%	0.2%
全体	462,367	139.9	400,517	98,953	75,472	37,308	19,577	1,862
			86.6%	21.4%	16.3%	8.1%	4.2%	0.4%

表中、「CN」は中国出願人からの出願（以下、「CN 出願」という。）、「CN 以外」は中国以外の諸外国からの出願（以下、「外国出願」という。）を指す。「JP」と「US」は、外国出願に含まれる日本からの出願（以下、「JP 出願」という。）と米国からの出願（以下、「US 出願」という。）の集計値である。

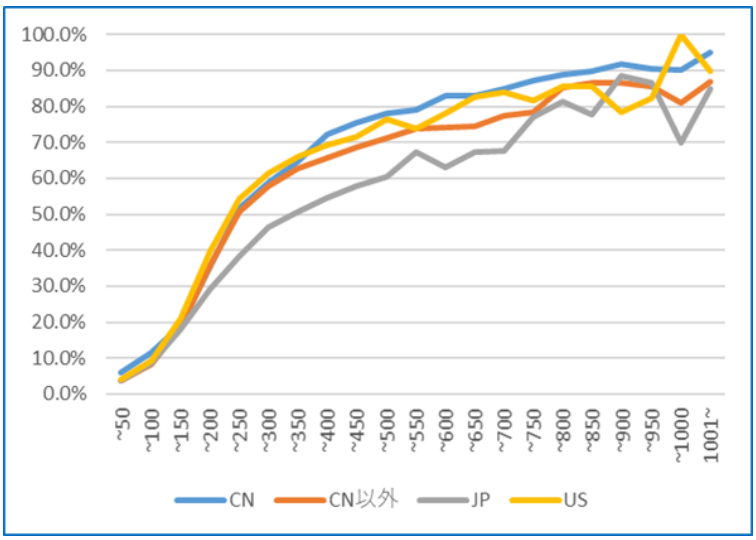
調査対象の全請求項において、CN 出願と外国出願とは 44.5%：55.5%の構成比率となった。それぞれの集計値を比較すると、CN 出願は一請求項あたりの平均文長が 177.6 文字と、外国出願（109.8 文字）に比べて極端に長い。

各種重大不備の発生率も CN 出願のほうが大幅に高いが、これは、機械翻訳精度が文長に相関することを考えれば、順当な結果である。CN 出願の平均文長 177.6 文字は、前項に示した IPC セクション別分析で最長であった E セクションの文長（177.7 文字）に匹敵する。各種不備の発生率もおおむね E セクションの数値と同水準であり、文長と重大不備の発生率との密接な相関性がここでも示された。

他方、外国出願に関しては、中国語が不自然で理解しにくいものが存在する懸念があった。これは原文の問題であり、機械翻訳処理の不備と判定すべきではないが、不自然な機械和訳は校正者が分かり易く校正するため、校正率には反映されるはずである。だが、実際の校正率（表中の「校正あり」）は 84.3%と、平均文長が 109.8 文字と短いことを考慮しても、特段に高い数値ではなかった。第 2.5.1 項に示した文長別統計における平均文長 50～100 文字の範囲の校正率 86.8%を下回っており、校正率はむしろ低いといえる。この結果から、少なくとも特許登録された外国出願であれば、CN 出願に比べて原文の品質が極端に悪いという懸念は少ないことがわかった。

事実、CN 出願と外国出願、さらに外国出願のうちの JP 出願と US 出願の重大不備発生率を文長範囲別に比較した結果（下図）においても、各文長範囲とも CN 出願と外国出願の間に顕著な差異は見られない。US 出願に関しても同様である。唯一、JP 出願のみは不備の発生率が概して低くなっているが、これは、JP 出願は中日対応特許として学習データでカバーされる確率が高く[29]、他よりも条件的に有利なためと考えられる。外国出願の不備発生率が各文長範囲で CN 出願よりも若干低くなっているのは、外国出願の約 15.0%を占める JP 出願の影響であろう。事実、下図を見ても、外国出願の約 33.3%を占める US 出願の当該範囲の不備発生率は CN 出願とほぼ同等である。

図 2.7-1. 国コード（CN／CN 以外／JP／US）×文長別 重大不備発生率（事業後期分）



[29] 中日対応特許は、ファミリーの中に中国特許出願と日本特許出願が含まれていることが条件となる（なお本編で「CN 出願」「外国出願（JP／US 含む）」と称したものは全て中国特許出願であり、それを出願人の国コードで区別している）。日本の出願人による中国特許出願（つまり JP 出願）は母国である日本国内でも出願される確率が高く、したがって CN 出願や他の外国出願よりも中日対応特許が存在する可能性が高いと考えられる。

次に、CN 出願の平均文長が外国出願よりも極端に長い理由を調べるため、それぞれの全請求項に占める独立請求項、従属請求項の構成比と重大不備の発生率を比較した。下表に結果を示す。

表 2.7-2. 出願人国コード別 独立請求項の校正有無と重大不備の発生率（事業後期分）

国	請求項数	平均文長	校正あり	重大不備	（重大不備の類型別内訳）			
					誤訳	訳抜け	湧出し	再翻訳
CN	36,214 (17.6%)	510.1	36,027	26,642	19,983	15,897	8,018	736
			99.5%	73.6%	55.2%	43.9%	22.1%	2.0%
CN 以外	32,385 (12.8%)	337.1	32,758	19,700	14,345	8,702	5,082	422
			99.8%	60.0%	43.7%	26.5%	15.5%	1.3%
全体	69,049 (14.9%)	427.9	68,785	46,342	34,328	24,599	13,100	1,158
			99.6%	67.1%	49.7%	35.6%	19.0%	1.7%

表 2.7-3. 出願人国コード別 従属請求項の校正有無と重大不備の発生率（事業後期分）

国	請求項数	平均文長	校正あり	重大不備	（重大不備の類型別内訳）			
					誤訳	訳抜け	湧出し	再翻訳
CN	169,571 (82.4%)	106.5	148,197	30,125	23,722	8,092	3,528	524
			87.4%	17.8%	14.0%	4.8%	2.1%	0.3%
CN 以外	2233,747 (87.2%)	76.4	183,535	22,486	17,422	4,617	2,949	180
			82.0%	10.0%	7.8%	2.1%	1.3%	0.1%
全体	393,318 (85.1%)	89.4	331,732	52,611	41,144	12,709	6,477	704
			84.3%	13.4%	10.5%	3.2%	1.6%	0.2%

上表のとおり、CN 出願は独立請求項が 36,214 項で、全請求項の 17.6%を占めている。一方、外国出願の独立請求項の構成比は 12.8%（32,385 項）であり、比率としては相対的にかなり少ない。この結果は、中国国内からの特許出願が諸外国からの特許出願に比べて一出願あたりの独立請求項の構成比が高いことを示している。また、文長に関しても、CN 出願の独立請求項の平均文長は 510.1 文字、従属請求項の平均文長は 106.5 文字と、ともに諸外国出願（337.1 文字、76.4 文字）を大きく上回っている。

つまり、中国国内からの特許出願は、理由は定かでないが、諸外国からの出願に比べて一出願あたりの従属請求項数が少なく、かつ一請求項あたりの文長が長くなる傾向があるといえる。その結果、各種重大不備の発生率が相対的に高くなっている。

2.8. 出願人の法人／個人別の集計結果と分析

続いて、出願人を法人と個人とに区分して集計した。中国特許公報には法人／個人を示す書誌項目は存在しないため、法人／個人の識別は出願人名称の表記（例えば「公司」を含む等）に基づいて行った。下表に、事業後期分の集計結果を示す。

表 2.8-1. 出願人の法人／個人別 校正と重大不備の発生率（事業後期分）

人	請求項数	平均文長	校正あり	重大不備	（重大不備の類型別内訳）			
					誤訳	訳抜け	湧出し	再翻訳
法人	449,945 (97.3%)	138.9	389,586	95,391	72,675	35,708	18,878	1,818
			86.6%	21.2%	16.2%	7.9%	4.2%	0.4%
個人	12,422 (2.7%)	177.7	10,931	3,562	2,797	1,600	699	44
			88.0%	28.7%	22.5%	12.9%	5.6%	0.4%
全体	462,367	139.9	400,517	98,953	75,472	37,308	19,577	1,862
			86.6%	21.4%	16.3%	8.1%	4.2%	0.4%

調査対象とする中国特許公報を選定する際、出願人が法人か個人かは考慮せず、無作為選定とした。結果、全体の 97.3%を法人出願が占め、個人出願は 2.7%、12,422 項と少数にとどまった。

このため、法人出願の集計結果は、全請求項の集計値とほぼ同値となっている。一方、個人出願の集計結果を見ると、原文の平均文長が顕著に長く、これにより各種不備の発生率も高くなっている。

原文の平均文長が長大となる傾向は、前項に示した国コード CN の出願にも見られた。そこで個人出願における CN 出願の構成比を調査したところ、全体の 66.8%（8,296 項）を占めており、法人出願における CN 出願の比率（43.0%、197,489 項）より顕著に高いことが判明した。個人出願は法人出願に比べて自国民の比率が高くなるのは自然であり、このことにより、個人出願の集計結果にも先述した国コード CN の傾向が強く反映されたと考えられる。とはいえ、中国特許に占める個人出願の比率は非常に低く、機械翻訳プラットフォーム全体の翻訳精度への影響も小さいと見なせる。

2.9. 英数字記号文字列を含む請求項の傾向

機械翻訳プラットフォームにおいて、原文に含まれる英数字記号文字列が翻訳精度に及ぼす影響の有無を統計的に調査した。

英数字記号自体は、請求項番号や句点、カンマ等を含めれば、全ての請求項に含まれている。ただし、長大な英数字記号文字列が含まれる請求項は、他の請求項に比べて翻訳精度に差異が生じている可能性が考えられる。このため、以下の各基準で集計を実施した。

・10 文字以上の連続する英数字記号文字列を含む請求項
・所定の条件に合致する数式を含む請求項
・所定の条件に合致する化合物名称を含む請求項

以下、それぞれの結果を示す。

2.9.1. 10 文字以上の連続する英数字記号文字列を含む請求項

英数字記号文字列は短いものまで含めれば全ての請求項に含まれているため、分析に際して何らかの基準を設ける必要がある。本調査では、10 文字以上の連続する英数字記号文字列の有無を基準として、これを含む請求項と、含まない請求項とを比較した。以下、結果を示す。

表 2.9.1-1. 10 文字以上の英数字記号文字列を含む請求項の重大不備発生率

項目		含む	含まない
請求項数	全	62,346 項（6.6%）	887,536 項（93.4%）
	後	30,507 項（6.6%）	431,860 項（93.4%）
一請求項あたりの平均文長	全	264.7 文字	129.5 文字
	後	264.6 文字	131.1 文字
一請求項あたりの平均改行数	全	3.3 回	2.0 回
	後	3.3 回	2.0 回
校正作業が発生した請求項	全	56,564 項（90.7%）	764,064 項（86.1%）
	後	27,797 項（91.1%）	372,720 項（86.3%）
重大な不備を含む請求項	全	19,423 項（31.2%）	162,140 項（18.3%）
	後	10,662 項（34.9%）	86,291 項（20.4%）
・重大な誤訳	全	14,551 項（23.3%）	122,481 項（13.8%）
	後	8,135 項（26.7%）	67,337 項（15.6%）
・重大な訳抜け	全	8,845 項（14.2%）	59,425 項（6.7%）

		後	4,748 項（15.6%）	32,560 項（7.5%）
	・重大な湧き出し	全	4,997 項（8.0%）	36,902 項（4.2%）
		後	2,279 項（7.5%）	17,298 項（4.0%）
	・再機械翻訳を使用	全	680 項（1.1%）	4,874 項（0.5%）
		後	213 項（0.7%）	1,649 項（0.4%）

まず、連続 10 文字以上の英数字記号文字列を含む請求項は、全期間を通じて全請求項の 6.6%と比較的少数であった。平均文長は 264.7 文字であり、それ以外の請求項の平均文長 129.5 文字の約 2.0 倍となっている。事業後期分のみの集計値も状況はほぼ同様であり、長大な英数字記号文字列を含む請求項は概して文長も長くなる傾向が見られる。

重大不備の発生率（事業後期分）を比較すると、10 文字以上の英数字記号文字列を含む請求項での不備発生率は 34.9%であり、含まない請求項の 20.4%よりも 14.5 ポイント高い。ただし、これは主に平均文長の差によるものと考えられる。それぞれの平均文長 264.6 文字、131.1 文字を第 2.5 項で調査した文長別の不備発生率と対比すると、前者は該当する文長範囲「201～300 文字」の 54.6%よりも顕著に低率であり、後者は「101～200 文字」の 24.0%から見ておおむね妥当な比率である。

より精密に検証を行うため、10 文字以上の英数字記号文字列を含む請求項の平均文長 264.6 文字に鑑み、文長を 201～300 文字の範囲に限定して双方を比較した。その結果を下表に示す。

表 2.9.1-2. 10 文字以上の英数字記号文字列を含む請求項の不備発生率（201～300 文字）

項目		含む	含まない
請求項数	全	7,253 項（11.9%）	53,643 項（88.1%）
	後	3,573 項（12.1%）	26,037 項（87.9%）
一請求項あたりの平均文長	全	245.5 文字	246.2 文字
	後	246.1 文字	246.1 文字
一請求項あたりの平均改行数	全	3.2 回	4.0 回
	後	3.2 回	3.9 回
校正作業が発生した請求項	全	7,145 項（98.5%）	53,367 項（99.5%）
	後	3,522 項（98.6%）	25,919 項（99.5%）
重大な不備を含む請求項	全	3,029 項（41.8%）	26,949 項（50.2%）
	後	1,653 項（46.3%）	14,416 項（55.4%）
・重大な誤訳	全	2,186 項（30.1%）	19,020 項（35.5%）

		後	1,209 項（33.8%）	10,078 項（38.7%）
	・重大な訳抜け	全	1,219 項（16.8%）	12,559 項（23.4%）
		後	621 項（17.4%）	6,696 項（25.7%）
	・重大な湧き出し	全	704 項（9.7%）	7,222 項（13.5%）
		後	305 項（8.5%）	3,294 項（12.7%）
	・再機械翻訳を使用	全	75 項（1.0%）	954 項（1.8%）
		後	28 項（0.8%）	318 項（1.2%）

表に示したとおり、201～300 文字の文長範囲において、10 文字以上の英数字記号文字列を含む請求項の重大不備発生率（事業後期分）は 46.3%であり、含まない請求項の発生率 55.4%よりも 9.1 ポイント低率であった。双方の平均文長はともに 246.1 文字で、文長による有利不利はない。この結果から、英数字記号文字列は機械翻訳プラットフォームの翻訳精度への特段の悪影響は有さず、むしろ有利に働いている可能性が高いと結論できる。

英数字記号は原則として原文文字列をそのまま機械和訳に転記すればよく、翻訳難易度は低い。このため、同じ文長であれば、英数字記号を多く含む請求項のほうが翻訳精度の面で有利となることは妥当といえる。ただし、これは機械翻訳プラットフォームのエンジンが英数字記号を適切に処理していることが前提となる。本調査結果は、その証左といえる。

2.9.2. 数式／化合物名称を含む請求項

前項の調査により、機械翻訳プラットフォームにおいて原文中の英数字記号文字列が不備の主な要因となっていないことが確認された。ただし、英数字記号文字列には、単純な数値の列挙から英名・ラテン語名の表記、数式や化学記号など多くの種類が存在し、特定の種類で不備が多発している可能性は皆無ではない。そこで、本調査では、英字・数字・記号が混在して構造が比較的複雑になる傾向が強い「数式」と「化合物名」について、所定の条件を設けて統計を行い、不備の発生状況を調査した。以下、その結果を示す。

2.9.2.1. 10 文字以上の数式を含む請求項

構造が複雑な数式の条件として、英数字記号 10 文字以上の文字列であって、数字を含んでおり、かつ数式に頻出の記号[30]を含み、さらに<sub>（下付文字タグ）を含まないもの[31]を選別した。以下、その集計結果を示す。

表 2.9.2.1-1. 10 文字以上の数式を含む請求項の重大不備発生率

項目		数式を含む	数式を含まない
請求項数	全	4,462 項（0.5%）	945,420 項（99.5%）
	後	2,301 項（0.5%）	460,066 項（99.5%）
一請求項あたりの平均文長	全	290.2 文字	137.7 文字
	後	292.2 文字	139.2 文字
一請求項あたりの平均改行数	全	4.0 回	2.1 回
	後	4.0 回	2.1 回
校正作業が発生した請求項	全	4,033 項（90.4%）	816,595 項（86.4%）
	後	2,091 項（90.9%）	398,426 項（86.6%）
重大な不備を含む請求項	全	1,427 項（32.0%）	180,136 項（19.1%）
	後	789 項（34.3%）	98,164 項（21.3%）
・重大な誤訳	全	1,074 項（24.1%）	135,958 項（14.4%）
	後	591 項（25.7%）	74,881 項（16.3%）
・重大な訳抜け	全	730 項（16.4%）	67,540 項（7.1%）
	後	406 項（17.6%）	36,902 項（8.0%）
・重大な湧き出し	全	389 項（8.7%）	41,510 項（4.4%）
	後	199 項（8.6%）	19,378 項（4.2%）
・再機械翻訳を使用	全	46 項（1.0%）	5,508 項（0.6%）
	後	18 項（0.8%）	1,844 項（0.4%）

10 文字以上の数式を含む請求項は、全請求項の 0.5%とごく少数であった。平均文長は 290.2 文字であり、これは前項に示した 10 文字以上の英数字文字列を含む請求項の平均 264.7 文字よりも 25.5 文字長く、長大化の傾向がさらに顕著に表れている。

一方、重大な不備を含む請求項の比率（事業後期分）は 34.3%と、10 文字以上の英数字文字列の平均 34.9%よりも 0.6 ポイント低くなった。僅差であり同水準と見るのが妥当だが、平均文長がより長いことを考えれば、数式を含む請求項の不備発生頻度は英数字文字

[30] 具体的には、＋、×、÷、＝、≠、>、<、≤、≥、≦、≧、*、・とした。

[31] <sub>は化合物名に多用され、ノイズとなるため。化合物名は別途調査している。

列全般よりもさらに低く抑えられているともいえる。

数式に関しても、ボリュームゾーンである文長範囲 201～300 文字の請求項に限定して不備発生率への影響を精査した。下表に示す。

表 2.9.2.1-2. 10 文字以上の数式を含む請求項の不備発生率（201～300 文字）

項目		数式を含む	数式を含まない
請求項数	全	454 項（0.7%）	60,442 項（99.3%）
	後	230 項（0.8%）	29,380 項（99.2%）
一請求項あたりの平均文長	全	245.9 文字	246.1 文字
	後	246.5 文字	246.1 文字
一請求項あたりの平均改行数	全	3.5 回	3.9 回
	後	3.6 回	3.8 回
校正作業が発生した請求項	全	449 項（98.9%）	60,063 項（99.4%）
	後	229 項（99.6%）	29,212 項（99.4%）
重大な不備を含む請求項	全	159 項（35.0%）	29,819 項（49.3%）
	後	85 項（37.0%）	15,984 項（54.4%）
・重大な誤訳	全	114 項（25.1%）	21,092 項（34.9%）
	後	62 項（27.0%）	11,225 項（38.2%）
・重大な訳抜け	全	66 項（14.5%）	13,712 項（22.7%）
	後	29 項（12.6%）	7,288 項（24.8%）
・重大な湧き出し	全	27 項（5.9%）	7,899 項（13.1%）
	後	12 項（5.2%）	3,587 項（12.2%）
・再機械翻訳を使用	全	2 項（0.4%）	1,027 項（1.7%）
	後	1 項（0.4%）	345 項（1.2%）

上表のとおり、201～300 文字の文長範囲では、数式を含む請求項の重大不備発生率は 37.0%（事業後期分）と、10 文字以上の英数字記号文字列を含む請求項の 46.3%よりも 9.3 ポイント低くなり、数式を含まない請求項（54.4%）との差は-17.4 ポイントと大きく開いた。対象となる請求項が 230 項と非常に少ないため極端な数値となっている可能性は否めないが、少なくとも上記検証結果により、数式も英数字文字列全般と同様、重大不備を増加させる性質ではなく、文長が同等であれば翻訳品質面でむしろ有利に働くものとわかる。

2.9.2.2. 英数字で表記された化合物名称を含む請求項

続いて、英数字で表記された化合物名称を含む請求項における不備の発生状況を確認した。対象とした請求項は、日本版ウィキペディアに掲載される化合物一覧[32]のうち、下付き数字を含む化合物名称を含んでいるものとした[33]。なお、文字数に関しては条件を設けておらず、10 文字以上の英数字記号文字列に該当しない化合物名称も集計対象に含まれる。以下、結果を示す。

表 2.9.2.2-1. 所定の化合物名称を含む請求項の重大不備発生率

項目		化合物名称を含む	化合物名称を含まない
請求項数	全	9,282 項（1.0%）	940,600 項（99.0%）
	後	4,614 項（1.0%）	457,753 項（99.0%）
一請求項あたりの平均文長	全	352.2 文字	136.3 文字
	後	354.1 文字	137.8 文字
一請求項あたりの平均改行数	全	3.8 回	2.1 回
	後	3.8 回	2.1 回
校正作業が発生した請求項	全	8,369 項（90.2%）	812,259 項（86.4%）
	後	4,156 項（90.1%）	396,361 項（86.6%）
重大な不備を含む請求項	全	3,180 項（34.3%）	178,383 項（19.0%）
	後	1,764 項（38.2%）	97,189 項（21.2%）
・重大な誤訳	全	2,281 項（24.6%）	134,751 項（14.3%）
	後	1,307 項（28.3%）	74,165 項（16.2%）
・重大な訳抜け	全	1,619 項（17.4%）	66,651 項（7.1%）
	後	874 項（18.9%）	36,434 項（8.0%）
・重大な湧き出し	全	879 項（9.5%）	41,020 項（4.4%）
	後	414 項（9.0%）	19,163 項（4.2%）
・再機械翻訳を使用	全	122 項（1.3%）	5,432 項（0.6%）
	後	42 項（0.9%）	1,820 項（0.4%）

所定の条件に合致する化合物名称を含む請求項は 9,810 項、全体 1.0%ほどであり、数式よりは多いものの、こちらもごく少数であった。平均文長 352.2 文字は数式（290.2 文

[32] https://ja.wikipedia.org/wiki/%E5%8C%96%E5%90%88%E7%89%A9%E4%B8%80%E8%A6%A7。日本語版ウィキペディアに記事が存在する化合物の一覧である。

[33] 下付き数字を含まない化合物名称（例：KI…ヨウ化カリウム）はノイズが多くなるため除外した。結果、英数字記号の化合物名称 621 種のうち 54 種が除外され、567 種の化合物名称が対象となった。

字）よりもさらに 62.0 文字長く、重大不備発生率（事業後期分）も 38.2%で、数式の 34.3%より 3.9 ポイント増加した。

化合物名称の場合、数式に比べて上付下付文字を表示する<sup>タグや<sub>タグが多用される。これらが文字数にカウントされたことが、数式よりも文字数が増加した理由と考えられる。事実、タグを文字数カウントから外せば、化合物名称を含む請求項は 285.5 文字となり、数式を含む請求項の平均文長 272.9 文字とほぼ同等である。

本調査の結果では、化合物名称を含む請求項は数式を含む請求項よりも重大不備発生率が 3.9 ポイント高かった。だが、62.0 文字という文長差も考慮すれば、数式に比べて特段に不備が多発しているとは言えない。つまり、化合物名称における<sup>タグや<sub>タグの多用は、（文長の長大化を除けば）機械翻訳精度に悪影響を及ぼしていない可能性が高い[34]。

化合物名称についても英数字記号文字列、数式と同様の検証を行った。ただし、文長範囲は平均文長 354.1 文字に鑑み、301～400 文字に設定した。下表に結果を示す。

表 2.9.2.2-2. 所定の化合物名称を含む請求項の重大不備発生率（301～400 文字）

項目		化合物名称を含む	化合物名称を含まない
請求項数	全	958 項（2.5%）	37,675 項（97.5%）
	後	430 項（2.3%）	18,231 項（97.7%）
一請求項あたりの平均文長	全	349.0 文字	346.8 文字
	後	346.7 文字	346.8 文字
一請求項あたりの平均改行数	全	4.0 回	5.4 回
	後	3.9 回	5.3 回
校正作業が発生した請求項	全	940 項（98.1%）	37,588 項（99.8%）
	後	425 項（98.8%）	18,199 項（99.8%）
重大な不備を含む請求項	全	443 項（46.2%）	22,223 項（59.0%）
	後	230 項（53.5%）	12,125 項（66.5%）
・重大な誤訳	全	312 項（32.6%）	16,173 項（42.9%）
	後	162 項（37.7%）	8,887 項（48.7%）
・重大な訳抜け	全	211 項（22.0%）	10,926 項（29.0%）
	後	105 項（24.4%）	6,007 項（32.9%）

[34] 現行の機械翻訳プラットフォームでは、上付下付タグは機械和訳に一切反映されていない。これは反映の失敗ではなく、仕様によるものと考えられる。

	・重大な湧き出し	全	111 項（11.6%）	6,850 項（18.2%）
		後	44 項（10.2%）	3,302 項（18.2%）
	・再機械翻訳を使用	全	11 項（1.1%）	879 項（2.3%）
		後	4 項（0.9%）	295 項（1.6%）

上表のとおり、化合物名称を含む請求項に関しても、数式や英数字記号文字列全般と同様、同程度の文長であれば重大不備発生率はむしろ低率であった。この結果から、種別を問わず、長大な英数字記号文字列の存在は、文長基準での翻訳精度比較において非常に有利に働くことがわかる。同時に、<sup>タグや<sub>タグの多用が機械翻訳精度に悪影響を及ぼしていないことも裏付けられた。

本項の結論としては、数式、英数字による化合物名称とも、英数字記号文字列の全体傾向がそのまま顕れており、機械翻訳プラットフォームにおいて、英数字記号文字列の翻訳に種別ごとの得手不得手は存在しないことが確認できた。機械翻訳のメカニズムの観点からも、英数字記号文字列は基本的には原文のまま転記すればよく、種別による難易度の差は生じにくいと考えられる。

なお、本項で調査した数式や化合物名称は英数字記号のみで構成されたものであるが、数式や化合物名称には中国語を含むものも存在する。また、化合物名称は中国語で記載されることも多い。こうしたケースにおける不備の状況に関しては、第 3 章（第 3.6 項、第 3.5.3 項）でさらに分析する。

3. 機械和訳における特徴的な不備の調査・分析結果

前章では、機械翻訳プラットフォームの中国特許公報機械和訳における各種不備の発生頻度を統計的に分析して、その傾向を把握した。本章では、校正作業を通じて収集した現状の機械和訳における特徴的な不備を、カテゴリ別に分析・考察する。

特徴的な不備は、特定の校正者が通年の校正作業（約 1,300 案件）を通じて検出した多数の不備の中から、発生頻度の多いものや翻訳精度への悪影響が大きいものを選定し、類型化した。その後、不備の各類型のおおよその発生頻度を把握するため、160 案件のサンプル（各セクション 20 件ずつ）を用いて、各類型の発生頻度を詳細に記録した[35]。以下、調査分析の結果を類型別に示す。

3.1. ニューラル機械翻訳方式の特徴

機械翻訳プラットフォームは、ニューラル機械翻訳方式のエンジンを使用している。ニューラル機械翻訳方式は、エンジンにあらかじめ多量の学習データを学習させ、学習結果に基づいて最も確からしい翻訳文を生成する。

学習データには、原語と翻訳語による対訳文データ（対訳コーパス）が使用される。まず、この学習データをデフォルト状態のエンジンに学習させる。具体的には、対訳コーパスの各文をトークン（翻訳の最小単位）に分解し、トークン単位の対訳語データを生成するとともに、文中の他のトークンとの関係性を示すベクトル値を与える。そして、このベクトル値に基づいて学習データの各原文に対して機械翻訳を試行し、その結果を学習データの対訳文と比較して、学習データ全体の翻訳精度が最良となるようにベクトル値を調整していく。この処理を繰り返し、学習データ全体でのベクトル値が最適化された時点で、エンジンの学習が完了する。

学習が完了したエンジンに翻訳対象文が入力されると、エンジンは入力文をトークンに分解し、文中の周辺語の状況に基づいてベクトル値を付与する。そして、各トークンのベクトル値に近似する学習データ中の訳語候補をいくつか選定し、複数の翻訳文候補を生成して、最も確からしい文を採用する。このようにして翻訳処理が行われる。

この方式により、学習データ中に複数の訳語候補が存在しても、入力文の内容（すなわち周辺の語の状況）に応じて、その都度最適と判断される訳語が採用される。反面、学習

[35] 詳細は添付資料 02『サンプル 160 件における特徴的不備の検出記録』、及び第 3.8 項に示したカウント結果一覧を参照されたい。なお、サンプル 160 件中、本章にまとめた各類型に該当する不備が一切検出されなかった案件は 160 件中 1 件（サンプル H14）のみであった。

データ中に存在しない用語は翻訳できず、訳語候補が存在しても、入力文のトークンと十分に近似なベクトル値でなければ（すなわち、周辺の語の状況がある程度類似していなければ）採用されない。

このように、ニューラル機械翻訳の翻訳精度は学習データの内容に大きく依存する。良好な翻訳精度を得るには、翻訳対象の文書と同種の対訳コーパスを大量に用意する必要がある。翻訳対象が特許文献である場合、学習データには主に対応特許公報由来の対訳コーパスが用いられる。

最後に、機械翻訳の処理単位について述べる。機械翻訳エンジンが 1 回の処理で取り扱える文字数（トークン数）には上限がある。このため、機械翻訳では、入力された文字列を句点（。）や改行で文単位に区切る前処理を行い、一文ずつ翻訳処理を行っている。また、請求項で多用されるセミコロン（；）も、一文が長大な場合に適度な文長に分割する区切り位置として扱われることが多い[36]。

なお、現状の機械翻訳では、翻訳処理単位間の情報の授受は未だ研究段階にあり、一般には実用化されていない。翻訳処理は個々の翻訳処理単位で完結し、先行する文の翻訳結果を記憶して後続文の訳語や表現を統一するといった文間の情報の共有は現時点では行われていない。

機械翻訳プラットフォームのエンジンや前後処理、学習データ等の詳細は非公開だが、機械和訳の状況から、上記特徴の全てに該当すると考えられる。本章では、これらニューラル機械翻訳方式の特徴を前提に、同プラットフォームの機械和訳に見られる各種不備を分析する。

[36] セミコロンで区切らずとも一括で翻訳処理できるような短文や、区切ると翻訳単位が非常に短くなるような場合は、区切らずに翻訳処理するのが一般的である。

3.2. 用語の誤訳

ニューラル機械翻訳方式における用語の誤訳は、学習データ中に存在しない語が誤訳されるケースと、学習データ中に適訳語が存在するにもかかわらず誤訳されるケースとに大別される。

3.2.1. 学習データ中に存在しない語の誤訳

ニューラル機械翻訳は、入力文を用語（トークン）単位に分解し、学習データ中に存在する語彙の中から、最も確からしい文を構成する訳語を採用する。このため、学習データ中に訳語が存在しなければ、正しく訳せない。該当する事例を示す。

［事例 3.2.1-1：CN107124391B ｾｸｼｮﾝ H］

原　　文	根据所述验证信息识别刷单行为，包括：识别所述无线接入点名称是否为随机字符串，若所述无线接入点名称为随机字符串，则确定所述客户端执行刷单行为；
機械和訳	前記検証情報に基づいてブラシ単体行動を識別し,以下を含む:前記無線アクセスポイント名がランダム文字列であるかどうかを識別し,前記無線アクセスポイント名がランダム文字列と呼ばれる場合,前記クライアントがブラシシング行動を実行するかどうかを決定する;
校正結果	前記認証情報に基づいて架空注文行動を識別することは,以下を含む:前記無線アクセスポイント名がランダム文字列であるか否かを識別し,前記無線アクセスポイント名がランダム文字列である場合,前記クライアントが架空注文行動を実行したと判断する;

本例原文の「刷单行为」は、ネット販売サイトで自社商品への架空の注文を（ロボットアクセスで）大量に行い、ランキングを向上させる行為を指す。近年使用され始めた新語である。訳語は校正結果の「架空注文行動」等が妥当だが、機械和訳では「ブラシ単体行動」「ブラシシング行動」と大きく誤訳されている。

「刷单行为／刷单」という語は、過去の中日対応特許公報では使用例がなく[37]、機械翻訳プラットフォームの学習データ中にも存在しないと考えられる。学習データに存在しない用語は、出現の都度、必ず誤訳される。「刷单行为」は本案件（CN107124391B）の請求項中に計 8 回出現するが、訳語は上掲の「ブラシ単体行動」「ブラシシング行動」各 1 回、「ブラシ列」× 4 回、「ブラシ動作」×1 回、そして訳抜け×1 回と、全て誤訳されて

[37] 当財団保有の、2021 年発行分までの中日対応特許のほぼ全件に由来する約 2.3 億文対の対訳コーパスを調査した結果である。以下同じ。

いる。

こうした誤訳は、「刷単行为／刷单」という語が学習データ中に存在しないために、エンジンが入力文を解析した際「刷单行为」または「刷单」＋「行为」という適切な単位でトークン化できなかったことによる。その結果、「刷」＋「单」＋「行为」と過度に分解された状態で翻訳されている。「刷」には「ブラシ」の意味があり、「单」は「单体」や「シングル」と訳された学習データから訳語が採用された可能性が高い[38]。「行为」は一般語だが、「行動」「動作」と妥当に訳されたものがある一方で、「列」と誤訳されたものも多い。「刷单」が正しくトークン化されなかった影響で「行为」も「行」と「为」に分解され、「行」が「列」と誤訳されたと推察される。

このように、ニューラル機械翻訳では、学習データ中に訳語が存在しない用語は、さらに細かく（場合によっては文字単位のトークンに）分解して個別に翻訳し、それらを連結するという形での訳出が試みられる。複合語の多くはこの方式で翻訳されていると推測されるが、大半の複合語は既存の技術用語が連結されたものであるため、正しい単位でトークンに分解でき、それぞれ適訳語が取得できる。このため、妥当な訳語が得られることが多い。一例を示す。

［事例 3.2.1-2：CN112914750B ｾｸｼｮﾝ A］

原　　文	2.根据权利要求 1 所述的一种癫痫患者用辅助口部防护装置，其特征在于：所述下压板(1)与上压板(2)之间为活动连接，下压板(1)的下表面亦设计有牙套(3)。
機械和訳	2.請求項 1 に記載のてんかん患者用補助口部保護装置であって,その特徴は以下のとおりである:前記下押さえ板(1)と上押さえ板(2)との間が可動接続され,下押さえ板(1)の下面にもマウスピース(3)が設計される。
校正結果	2.請求項 1 に記載のてんかん患者用の補助的な口部保護装置であって,その特徴は以下のとおりである:前記下部押圧板(1)と上部押圧板(2)との間は可動接続され,下部押圧板(1)の下面にもマウスピース(3)が設計される。

本例の発明の主題である「辅助口部防护装置」は中日対応特許公報で使用例がなく、出願人の造語と見られる。機械和訳の「補助口部保護装置」は、原語を「辅助」「口部」「防护」「装置」という一般語に分解して直訳し、再連結したのみだが、それでも校正結果の

[38] 「刷」は「ブラシをかける、ブラッシングする」という動詞としても使われる。「ブラシシング」という誤訳は、「ブラッシング」が不正確な形で採用された結果の可能性もある。

「補助的な口部保護装置」とほぼ同じ理解が得られる。

これに対し、前事例の「刷単」のような新規の用語は、学習データ中にトークンが存在しないため、過度に分解され、正しく訳せない。こうした新語の誤訳は、学習データに訳例を追加することが唯一の解決策となる。新規の技術用語は、定着するにつれて特許公報での使用頻度も高くなる。したがって、最新の対応特許を定期的に学習させていけば、誤訳は高確率で解消する。

新語の中には、既存の用語を組み合わせて新たな概念を表すものもある。このタイプの新語も、学習データ中に訳例が存在しないことで誤訳される場合がある。一例を示す。

［事例 3.2.1-3：CN110927791B ｾｸｼｮﾝ G］

原　　文	7.根据权利要求 1 至 6 中任一项所述的方法，其特征在于，所述第一套深度学习网络和第二套深度学习网络基于卷积神经网络框架 Caffe 来构建，并采用 Python 编程语言来进行网络控制。
機械和訳	7.請求項 1 から 6 のいずれか一項に記載の方法であって,その特徴は以下のとおりである:前記第一セットの深さ学習ネットワーク及び第二セットの深さ学習ネットワークは畳み込みニューラルネットワークフレームワーク Caffe に基づいて構築され,且つ Python プログラミング言語を用いてネットワーク制御を行う。
校正結果	7.請求項 1 から 6 のいずれか一項に記載の方法であって,その特徴は以下のとおりである:前記第一セットのディープラーニングネットワーク及び第二セットのディープラーニングネットワークは,畳み込みニューラルネットワークフレームワーク Caffe に基づいて構築され,且つ Python プログラミング言語を用いてネットワーク制御を行う。

本例の「深度学习（ディープラーニング／深層学習）」は、近年、AI 分野の専門用語として定着した新語である。機械和訳はこの新語に対応できず、「深さ学習」と誤訳されている。

これは、学習データ中に「深度学习」を「ディープラーニング」又は「深層学習」と訳した訳例が存在せず、「深度」と「学习」に分解して個別に翻訳した結果と考える。現状、「深度学习」は調査対象のほぼ全ての請求項[39]で「深さ学習」又は「深度学習」と誤訳

[39] 「深度学习」を含む請求項 184 項のうち「深層学習」と適訳で訳されたものは 1 件のみ。大半は「深さ学習」と「深度学習」のいずれかで、例外は「奥行学習」が 1 件、訳抜けが 4 件のみであった。

されている。

このように、既存の用語を組み合わせた複合語であっても、新語として特定の訳語の組み合わせを用いなければならないものがある。このタイプの新語も、古い学習データから適訳語を得ることは難しい。本例も、「深度」＋「学習」という分解自体は正しい単位で行われており、かつ「深度→ディープ」、「学習→ラーニング」というトークンが学習データ中に存在する可能性も高い。だが、「深度学習→ディープラーニング」という訳例が学習されていない状況では、数ある訳語候補の中から「深度」と「学習」に対し的確に「ディープ」と「ラーニング」の組み合わせが採用されることは期待できない。「深度学習」が「ディープラーニング」（又は「深層学習」）と訳された最新の学習データを補充して、「深度」と「学習」が隣接する場合はそれぞれ「ディープ」「ラーニング」と訳すようエンジンに学習させる必要がある。

なお、学習データ中に存在しない用語は、新規の技術用語とは限らない。専門性が高い技術用語や出願人の造語など、使用頻度自体が低い語も、カバーされていない確率が高くなる。これに該当する事例を示す。

［事例 3.2.1-4：CN108265104B ｾｸｼｮﾝ C］

原　　文	4.根据权利要求 1 所述的构建方法，其特征在于，采用杜恩斯匀浆器对所述裂解产物进行松锤研磨不少于十次，紧锤研磨不少于十次，得到所述研磨产物。
機械和訳	4.請求項 1 に記載の構築方法であって,その特徴は以下のとおりである:ダンホモジナイザーを用いて前記溶解生成物に対して 10 回以上のハンマーミリングを行い,ハンマーミリングが 10 回以上行われ,前記研磨製品が得られる。
校正結果	4.請求項 1 に記載の構築方法であって,その特徴は以下のとおりである:ダウンス型ホモジナイザーを用いて前記溶解産物に対して粗調整用ペストルでの粉砕を 10 回以上行い,微調整用ペストルでの粉砕を 10 回以上行い,前記粉砕産物を得る。

本例は染色体関連の発明である。本例原文の「松锤研磨」と「紧锤研磨」は、それぞれ「粗調整用ペストルでの粉砕」「微調整用ペストルでの粉砕」という意味で使われているが、両語とも中日対応特許で使用例がなく、Google でも本案件以外ヒットしない。出願人の造語と見られ、機械翻訳プラットフォームの学習データでもカバーされていない可能性が高い。このため、本来は「松锤」＋「研磨」、「紧锤」＋「研磨」と分解されるべきところ、それぞれ「松」＋「锤研磨」、「紧」＋「锤研磨」と誤った分解がなされ、「锤研磨」

の部分のみが「ハンマーミリング」と訳出されたと推察される（「松[40]」と「紧」は「锤研磨」と組み合わされた使用例が過去になく、近似のベクトル値を持つトークンが存在しないため訳出不能となった可能性が高い)。発明内容からはハンマーを用いるとは考えにくく、誤訳と見なせる。

こうした語も、学習データ中に訳例が追加されれば、誤訳は解消する。だが、「刷单行为」や「深度学习」のような新語と異なり、こうしたマイナーな用語は今後の特許公報でも使用頻度が増加するとは考えにくい。したがって、学習データを定期的にアップデートしても、カバーされる確率は低い。誤訳を確実に解消するには、ピンポイントで対訳コーパスを作成し、学習データに加える必要がある。ただし、これを行うにはまず誤訳されているマイナーな用語を機械和訳から検出・収集しなければならない。かつ、こうした語は入力文での使用頻度も低いため、対策を講じても、全体の翻訳品質の改善効果は小さい。費用対効果の観点からの検討が必要である。

学習データで網羅することが困難な用語のジャンルとしては、人名や組織名、製品名といった固有名詞も挙げられる。ただし、特許請求項では固有名詞は著名な製品名称などを除き、ほぼ使用されない。明細書全文に範囲を広げても、固有名詞が発明の内容理解に深く関わるケースは限定的である。このため、特許公報の翻訳という観点からは、対応の優先度は低い。

サンプリングした 160 案件（A～H セクション各 20 件）のうち、学習データ中に訳語が存在しないためと見られる用語の誤訳は 44 案件（27.5%）で発生していた。その大半は事例 3.2.1-4 に例示したようなマイナーな技術用語の誤訳であったが、中には「大数据（ビッグデータ)」が「大データ」と誤訳された案件(サンプル E13)や、「充电桩（充電スタンド)」が「充電杭」「電気杭」「帯電杭」と誤訳された案件（サンプル E08）など、「深度学习（ディープラーニング)」と同様、近年一般的になった新規な技術用語（新語）が誤訳されるケースも存在した。この結果からも、機械翻訳プラットフォームの学習データで対応できていない新語の存在が示唆される。

[40] 松は他の樹木に比べて葉が細く、隙間が多く見えることから、「松」の字には「スカスカしている、疎である、ゆるい、粗い」といった意味がある。

3.2.2. 学習データ中に適訳語が存在する語の誤訳

ニューラル機械翻訳では、学習データ中に適訳語が存在しても、必ずその適訳語が採用されるとは限らない。主な原因として、用語が正しい単位で認識されないことと、訳語候補が複数存在し、入力文の分野や文脈に合わない訳語が採用されてしまうことが挙げられる。

［用語が正しい単位で認識されないことによる誤訳］

正しい機械翻訳文が出力されるには、まず入力文が各用語（トークン）に正しく分割されなければならない。用語の適訳語が学習データ中に存在していても、エンジンが入力文からその用語を正しい単位で分割できず、途中で切れたり、余計な文字が付随したりすると、学習データのトークンと合致せず、適切な訳語が得られない。以下、一例を示す。

［事例 3.2.2-1：CN113235177B ｾｸｼｮﾝ D］

原　　文	S3、通过侧吹风对复合纤维进行冷却成型，风温为 13-15℃，风速为 1-1.4m/s，风湿度为 80-90%,冷却吹风位置在喷丝板下方 40-60cm 处；
機械和訳	S3,サイドブローにより複合繊維を冷却成形し,風温は 13-15℃であり,風速は 1-1.4m/s であり,リウマチ度は 80-90%であり,冷却送風位置は紡糸口金の下方 40-60cm である;
校正結果	S3,サイドブローにより複合繊維を冷却成形し,風温は 13-15℃であり,風速は 1-1.4m/s であり,風の湿度は 80-90%であり,冷却送風位置は紡糸口金から 40-60cm 下の箇所である;

原文の「风湿度」は、校正結果にあるとおり、ここでは「風の湿度」を意味する。したがて「风」＋「湿度」と分解するのが正しい。「风」、「湿度」とも一般用語であり、学習データ中に「風」、「湿度」と訳された事例が多数存在すること（つまり、本例の適訳語が存在すること）は確実である。だが、実際の機械和訳では「リウマチ度」と大きく誤訳されている。

この誤訳は、「风湿」が中国語で「リウマチ」を指すことに起因する。つまり、エンジンが入力文中の「风湿度」を「风」＋「湿度」ではなく「风湿」＋「度」と分解したことによる誤訳である。「风湿度」という複合語自体は中日対応特許公報でも使用例がなく[41]、機械翻訳プラットフォームの学習データでもこの単位ではトークン化されていないと推察される。このため、さらなる分解が行われ、学習データ全体のベクトル値から「风」＋

[41] 「排风湿度」などで数文の使用例があるが、ここから得られるトークンは「排风湿度」又は「排风」＋「湿度」であり、「风湿度」というトークンは得られない。

「湿度」ではなく「风湿」＋「度」という解釈が優先されたと考えられる。この解釈が、本例では誤ったトークン分割となった。

上例は用語レベルの局所的な分割誤りだが、より広範囲の誤解釈、例えば周辺の語のトークン分割の失敗や、文全体の分割の失敗によって、本来は正しく翻訳できるはずの用語が突発的に誤訳されることもあり得る。例えば、前掲の事例 3.2.1-1 で「刷单行为」の影響により一般語の「行为（行動、行為）」がしばしば誤訳されたのも、このケースに該当する。ただし、このように用語が本来の単位で認識されず、意味の全く異なる訳語に誤訳されるケースは非常に限定的であり、機械翻訳プラットフォームでは大半の語が、文の構造によらず[42]、正確にトークン分割されていた。むしろ 160 件のサンプル調査では、複数の単語からなる複合語が、各単語レベルでは正しく認識されているものの、複合語として認識されず、個別に扱われて誤訳となるケースのほうが目立った。例えばサンプル A7 では「第一主动关节组件（第一能動関節アセンブリ）」が複合名詞と見なされずに細分化された結果、「第一能動関節アセンブリ上に設置され」と訳すべきところ「第 1 の能動的に関節部材上に設けられ」と誤訳されている。160 件中「用語が正しい単位で認識されないことによる誤訳」としてカウントされた 27 件の大半がこのケースに該当するものであった。

［訳語候補が複数存在することによる誤訳］

学習データ中に適訳語が存在するにもかかわらず誤訳が発生するもう一つの状況は、有力な訳語が複数存在する用語において、入力文の技術分野や文脈に合わない訳語が採用されてしまうケースである。

このケースでは、入力文のトークン分割は適切な単位で行われる。だが、そのトークンに対応する訳語候補が学習データ中に複数存在し、入力文にフィットしない訳語候補が採用されることで不適訳となる。これは、エンジンの学習の際に学習データ全体でベクトルを最適化した結果、適訳語よりも不適訳語のほうに、より近似のベクトル値が付されたためである。以下、一例を示す。

［事例 3.2.2-2：CN111945447B ｾｸｼｮﾝ D］

原　　文	4.根据权利要求 1 所述的一种基于聚丙烯酸型高吸水复合树脂的吸水纸的制备方法，其特征是：所述上网抄纸步骤中上网的浓度为 0.3-1wt%，浆料的打浆度为 28-42°，制得的湿面纤维层和湿底纤维层的含水率控制为 55-

[42] ただし長大な文を除く。長大な文は翻訳処理に不向きであり、用語の誤訳をはじめ訳抜け、湧き出しなど各種不備が多発する傾向が強い。

	70%。
機械和訳	4.請求項 1 に記載のポリアクリル酸系高吸水性複合樹脂に基づく吸水紙の製造方法であって,その特徴は以下のとおりである:前記ネットサーフィンステップにおけるインターネットの濃度は 0.3-1wt%であり,パルプの叩解度は 28-42°であり,製造された湿潤面繊維層と湿潤底繊維層の含水率は 55-70%に制御される。
校正結果	4.請求項 1 に記載のポリアクリル酸系高吸水性複合樹脂に基づく吸水紙の製造方法であって,その特徴は以下のとおりである:前記ワイヤ上抄紙ステップにおけるワイヤ上の濃度は 0.3-1wt%であり,スラリーの叩解度は 28-42°であり,得られる湿潤表面繊維層と湿潤底面繊維層の含水率は 55-70%に制御される。

本例原文の「上网」の「网」は「網、格子、ふるい」を意味する語だが、通信分野では「インターネット」の略称としても多用される。つまり、技術分野や発明内容によって適訳語が変わる語である。

人間がこうした語を読解する際は、発明の内容や文脈から、どちらの意味で使われているかを判断している。ニューラル機械翻訳では、この判断を、学習データ全体で最適化した各トークンのベクトル値の近似度（端的に言えば、周辺の語句の共通性）によって代行している。

本例は製紙工程に関する発明であり、「网」は製紙時に用いる網状の「ワイヤ[43]」を指しているが、機械和訳はこれを「インターネット」、「ネットサーフィン」と誤訳している。

本例の場合、「上网」というトークン単位自体は正しい[44]。誤訳の原因は、このトークンに複数の訳語候補が存在すること、そして全体最適化された学習データのベクトル値において、本例の「上网」のベクトル値に最も近似する訳語がそれぞれ「インターネット」、「ネットサーフィン」であったことによる。

インターネットに無関係な請求項において「上网」を「メッシュ（CN110938937B）」や「網（CN112411241B）」等の適訳語に訳しているケースも存在しており、「上网」が一律インターネット関連用語に誤訳されるわけではない。だが、全体的には誤訳が大多数で

[43] 製紙分野では、紙料を載せてシート状にするための網は「ワイヤ」と呼ばれる。

[44] 「上网」の「上」は「インターネットの」「ワイヤ上」の「の」「上」に相当する。

あり[45]、「上网」に関しては、学習データ中でインターネット関連用語が支配的な訳語となっている状況が窺える。ただし、全技術分野を俯瞰すれば「上网」は「ワイヤ」や「網」よりも「インターネット」の意味で用いられることが多く[46]、全体最適化という観点からは、理に適った訳語選定とも見なせる。

本項に示したように、学習データ中に適訳語が存在する用語の誤訳は、特定の文におけるトークン分解の失敗や、ベクトル値の全体最適化の結果に起因している。このため、学習データにどのような訳例を追加すれば（あるいは除去すれば）誤訳が解消されるか、予測が困難である。後者の場合、例えば「上网」の製紙分野の文例を集中的に学習させることで、ベクトルの全体最適化の結果を変化させ、「ワイヤ」や「網」が採用され易くすることはできる。しかしながら、その結果、それまで「インターネット」と適切に訳されていた多数の案件で「ワイヤ」や「網」が採用されて翻訳品質が悪化する懸念が大きく、推奨できない。対処が難しいタイプの誤訳といえる。

サンプリングした 160 案件（A〜H セクション各 20 件）において、複数の適訳語が存在することによる用語の誤訳は 19 件検出された。例えばサンプル G16 の「漫游」は VR 環境内を散策するという意味で使われているが、通信分野では「ローミング」の意味で多用されており、本案件でも「ローミング」と誤訳されている。また D2 では「打开」という動詞を「(モータを）オンにする」の意味で使用しているが、より一般的な「開放する」「開く」という訳語が採用され、「モータを開放する／開く」という、誤解を招く訳文となっている。

---

[45] 原文に「上网」と「纸」を含む（つまり製紙関係の発明と見られる）請求項 88 項中の 68 項で「上网」が「インターネット」又は「ネットサーフィン」と誤訳されていた。

[46] 中日対応特許公報で「上网」は 1,505 文に出現し、うち 618 文が「ネットワーク」、170 文が「インターネット」、109 文が「ネットブック」、22 文が「ネットサーフィン」と訳されている。「ウェブ」（56 文）も、その大半はインターネットの意味で用いられている。

3.2.3. 係り受けの誤訳

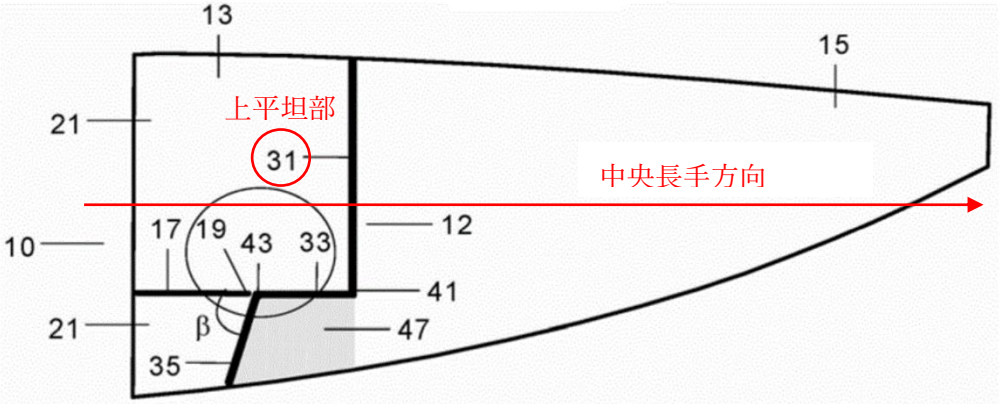
係り受けとは、語同士の文法的な関係性を指す。誤訳は、用語の誤訳と、係り受けの誤訳とに大別される。ただし、ニューラル機械翻訳方式においては、どちらの誤訳も不適切なトークンが採用された結果であり、発生のメカニズムは共通している。本項で、係り受けの誤訳について分析する。まず一例を示す。

［事例 3.2.3-1：CN105905272B ｾｸｼｮﾝ B］

原　　文	上平坦部(31)和下平坦部(35)，所述上平坦部(31)和所述下平坦部(35)形成与所述底板(17)的所述后边界(19)相邻接的弯折部(41)，所述上平坦部(31)与所述中心纵向轴线垂直，所述下平坦部(35)与所述底板(17)形成包括在 30° 与 60° 之间的角度；
機械和訳	上平坦部(31)と下平坦部(35)であって,前記上平坦部(31)と前記下平坦部(35)は前記底板(17)の前記後境界(19)に隣接する屈曲部(41)を形成し,前記上平坦部(31)と前記中央長手方向軸に垂直であり,前記下平坦部(35)と前記底板(17)は 30° と 60° の間に含まれる角度を形成する;
校正結果	上平坦部(31)と下平坦部(35)であって,前記上平坦部(31)と前記下平坦部(35)は前記底板(17)の前記後縁部(19)に隣接する屈曲部(41)を形成し,前記上平坦部(31)と前記中心縦軸は垂直であり,前記下平坦部(35)と前記底板(17)は 30° と 60° の間に含まれる角度を形成する;

本例原文の「所述上平坦部(31)与所述中心纵向轴线垂直」は、校正結果のとおり「前記上平坦部(31)と前記中心縦軸は垂直であり」という意味であるが（下図参照)、機械和訳では「前記上平坦部(31)と前記中央長手方向軸に垂直であり」となっており、不正確である。文法的にも無理がある解釈で、典型的な係り受けの誤訳である。

[参考] CN105905272B 図 5



ここで、ニューラル機械翻訳における係り受けの誤訳の発生メカニズムを説明する。前述のとおり、ニューラル機械翻訳は、入力文をトークンに分解し、その単位で学習データ中から最適の訳語を選択する。ただし、トークンの単位は学習データ中で頻出の「文字列」であり、必ずしも文法上の「語」とは限らない。語よりもさらに細かく、文法的には区切れない単位まで分割されることもあれば、複数の語が連なったフレーズ単位となることもある。以下、トークン分割の一例を示す。

[参考] 中日対訳文のトークン分割と対応づけ（アライメント）の実例

原　文	本发明涉及用作具有放射性核素螯合剂的放射性药物和放射成像剂的化合物。
対訳文	本発明は、放射性核種キレート剤を保有する放射性医薬品及び放射線造影剤として有用な化合物に関する。

↓↓↓

中国語	日本語
本发明涉及	本発明は／に関する[47]
用作	有用な
具有	を
放射	、／放射
性	性
核	核
素	種
螯	キ／レート
合	
剂的	剤
放射	放射
性	性
药物	医薬
和	品／及び
放射	放射線
成像	造／影
剂	剤として
的化合物	化合物
。	

[47] トークン分割は原文、対訳文それぞれ別個に行い、その後、双方のトークンを対応づける。上表「日本語」欄のスラッシュ（／）は、原文（中国語）の 1 トークンに対し、対訳文（日本語）の複数のトークンが対応付けられたことを示している。

	保 ／有する／。

上例では、例えば「剤」という中国語名詞に対し、日本語トークンは「剤として」と、助詞を伴う形で対応付けられている。中国語には「てにをは」に相当する助詞が存在しないため、このような対応づけがなされることが多い。この「剤：剤として」というトークンは、例文のように「用作：有用な」と係り受けがある場合には適訳となるが、あらゆる「剤」に対して使えるわけではない。学習データ中には、「剤」のトークンとして他にも「剤を」「剤に」「剤で」など様々な訳例が存在し、ニューラル機械翻訳は、周辺語句との関係性を表すベクトル値によって、その中から入力文に最適のトークンを都度採用する。つまり、ニューラル機械翻訳は、文法規則に基づいて翻訳文を構築するのではなく、正確な文法で書かれた大量の学習データに準拠することで、結果的に正しい文法の翻訳文を作成している。

前掲の事例 3.2.3-1 に関しても、中国語「轴线」に対して、「軸に」「軸は」をはじめ「軸が」「軸を」などさまざまな訳語のトークンが存在すると考えられる。どれが採用されるかは、文法規則ではなく、ベクトル値に基づく。このため、入力文の内容や学習データの状況によっては、訳語の選択を誤り、結果的に非文法的な訳文が生成される場合もあり得る。上例はこれに該当するケースと見なせる。

つまり、ニューラル機械翻訳における係り受けの誤訳は、学習データ中の訳語候補（トークン）から入力文に不適なものを採用した結果であり、本質的には用語の誤訳と同じ発生メカニズムといえる。

なお、上掲のトークン分割例からも明らかなように、中国語を日本語に翻訳する際には、語順（トークンの順序）も並べ替える必要がある。例えば中国語文の冒頭の「本发明涉及」は、日本語文に訳す際には文頭に「本発明は」、文末に「に関する」を配置しなければならない。ニューラル機械翻訳は、こうした翻訳文の語順決定にも文法規則を用いず、学習したベクトル値により各トークンをさまざまな訳語と語順に組み替えて複数の翻訳文候補を生成し、最も確からしいと判断した一文を採用する。文法に基づいて翻訳文を生成していないため、文法規則に反する語順が採用される可能性もある。以下、不正な語順決定により係り受けの誤訳が生じたと見受けられる事例を示す。

［事例 3.2.3-2：CN107790355B ｾｸｼｮﾝ B］

原　　文	基板收容部，包含可相互连接分离的第 1 覆盖部及第 2 覆盖部，将支撑于所述支撑部的所述基板收容于所述第 1 覆盖部与所述第 2 覆盖部之间的处理空间；

機械和訳	基板収容部であって,互いに分離可能に分離された第 1 カバー部及び第 2 カバー部を含み,前記支持部に支持された前記基板を前記第 1 カバー部と前記第 2 カバー部との間に収容する処理空間;
校正結果	基板収容部であって,互いに接続分離可能な第一被覆部及び第二被覆部を含み,前記支持部に支持された前記基板を前記第一被覆部と前記第二被覆部との間の処理空間に収容する;

本例は、原文の「收容于」→「に収容する」、「处理空间」→「処理空間」という訳語自体は適切なものが選択されているが、語順決定時に「処理空間」→「に収容する」ではなく「に収容する」→「処理空間」が選ばれ、誤訳となっている。文法的にはこの原文を機械和訳のように解釈するのは無理があるが、ニューラル機械翻訳は文法規則ではなく学習データから導出した確からしさで語順を決定しているため、こうした不自然な係り受け誤りが生じることもある。

このように、ニューラル機械翻訳における係り受け誤りは、用語の誤訳と同様、学習データの内容・状況に起因している。したがって、適切な学習データを追加することで改善される可能性がある。ただし、用語の誤訳に比べると、どのような学習データを追加すれば誤訳が改善されるかの予測が難しく、また、特殊な構造の文を過度に学習させると通常の文の翻訳品質に悪影響を及ぼす危険性もある。

係り受けの誤りはサンプリングした 160 案件中 84 件と過半数で発生していた。前掲の事例 3.2.3-1 に相当する「不適切な助詞による係り受け誤り」が 21 件、3.2.3-2 に相当する「不適切な語順による係り受け誤り」が 63 件であった。それぞれ典型例として、前者はサンプル A9 の請求項 1 で「濾過網が 1000 メッシュであるフィルタを用いて濾過し」が「濾過網を 1000 メッシュのフィルターで濾過し」と誤訳されたケース（語順は合っているが助詞が誤っている）、後者はサンプル A8 の請求項 1 で「前記出力ユニットによって決定されたクリーニングに関連する情報を視覚的に出力すること」が「前記出力ユニットによってクリーニングに関連する情報の視覚的出力を決定すること」と誤訳されたケース（「決定」の語順が誤っている）が挙げられる。

3.2.4. 事例集

本調査分析で検出した、用語の誤訳に該当する特徴的な事例を掲載する。

［事例 3.2.4-1：CN109943931B ｾｸｼｮﾝ D］

原　　文	清棉：抓棉打手转速为 700~900r/min，开棉机梳针滚筒转速为 300~500r/min，梳针滚筒与给棉罗拉之间的隔距 10~13mm，综合打手转速为 900~1000r/min，综合打手与天平之间隔距为 8~12mm，棉卷罗拉转速为 10~15r/min；
機械和訳	綿:綿とんすけの回転数は 700~900r/min であり,開綿機の針ドラムの回転数は 300~500r/min であり,針ドラムとフィードローラとの間の距離は 10～13mm であり,総合とんすけ回転数は 900~1000r/min であり,総合とんすけと天秤との間隔距離は 8～12mm であり,ロールローラの回転数は 10~15r/min である;
校正結果	打綿:プラッキングビーターの回転数は 700~900r/min であり,開綿機のキルシュナビーターの回転数は 300~500r/min であり,キルシュナビーターとフィードローラとの間の距離は 10～13mm であり,一体化ビーターの回転数は 900~1000r/min であり,一体化ビーターと天秤との間隔距離は 8～12mm であり,ラップローラの回転数は 10~15r/min である;
解　　説	機械和訳は全体的に不適訳語が目立つが、特に原文の「打手」が全て「とんすけ」と誤訳され、理解困難である。「とんすけ」は映画に登場するキャラクター名で、中国語名が「打手」であるが、特許文献では明らかに不適である。 サンプル 160 件では類似の事象は検出されなかったものの、10 万件全件では他にも「漂白箱（漂白タンク）」が「漂白 SHIROBAKO」（CN113512838B）、「大剣杆（大型レピア）」が「CLAYMORE ロッド」（CN112981671B）とそれぞれアニメの題名で訳された事例も検出されており、特許文献翻訳に不向きなエンターテインメント系の学習データが使用されている懸念がある。

［事例 3.2.4-2：CN109457541B ｾｸｼｮﾝ D］

原　　文	1.一种降焦烟用接装纸，其特征在于，…
	2.根据权利要求 1 所述的降焦烟用接装纸，…
	：
	10.根据权利要求 1 所述的降焦烟用接装纸，…
機械和訳	1.フォーカスタバコ用チッピング紙であって,その特徴は以下のとおりである:…

	2.請求項 1 に記載のタール用のチッピング用チッピング紙であって,…
	3.請求項 1 に記載のタール用チッピング用チッピング紙であって,…
	4.請求項 1 に記載のタール低減用チッピング紙であって,…
	5.請求項 4 に記載のタール用チッピング用チッピング紙であって,…
	6.請求項 1 に記載のタール用チッピング用チッピング紙であって,…
	7.請求項 1 に記載のタール用チッピング用チッピング紙であって,…
	8.請求項 1 に記載のタール用チッピング用チッピング紙であって,…
	9.請求項 1 に記載のタール用チッピング用チッピング紙であって,…
	10.請求項 1 に記載のタール低減用チッピング紙であって,…
校正結果	1.タール減少用チッピング紙であって,その特徴は以下のとおりである:…
	2.請求項 1 に記載のタール減少用チッピング紙であって,…
	:
	10.請求項 1 に記載のタール減少用チッピング紙であって,…
解　　説	本件は発明の主題「降焦烟用接装紙（タール減少用チッピング紙)」が全請求項に記載されている。「降焦烟」、「焦烟」は中日対応特許に使用例が存在せず、学習データ中にも訳例が存在しないと見られる。 学習データに存在しない語は常に誤訳されるという傾向に準じ、本例の機械和訳でも「降焦烟」は 10 回中 8 回「フォーカスタバコ」又は「タール用チッピング」と誤訳されている。 「フォーカスタバコ」は、「焦」から「フォーカス（焦点)」、「烟」から「タバコ（烟草)」を採用したが「降」が訳出できなかった結果と考えられる。一方、「タール用チッピング」に関しては、「タール」は「焦」に由来するが、「チッピング」に対応する語は「降焦烟」には存在しない。これは、訳出不可能な語の代わりに文中の別の語句を重複出力するという機械翻訳プラットフォームのエンジンの性質によるものと見られる。詳しくは第 3.4 項で後述する。 なお、10 回中 2 回のみは「タール低減」と正しく訳されたが、これは「降焦烟」が学習データ中に存在したからではなく、文字単位に分解して翻訳した結果、偶然に適切な訳語の組み合わせになったものと考えられる。

［事例 3.2.4-3：CN106579024B ｾｸｼｮﾝ A］

原　　文	7.如权利要求 1 所述的梯度微粉碎同步熟制骨粉和骨糊的加工方法，其特征在于，所述步骤四中还包括以下步骤：将初级骨粉制备成成品骨粉：经过离心喷雾后的初级骨粉再经过行星球磨机研磨至粒径为 10～12μm 的超细微的成品骨粉，其中，行星球磨机的转速为：450～500r/min，研磨时间为 3～4h；之后，内层用洁净塑料袋，每袋按 1Kg/袋进行包装得到骨粉产品，密闭入库，在干燥处阴凉处保存。
機械和訳	7.請求項 1 に記載の勾配微粉砕同期熟成骨粉と骨ペーストの加工方法であって,その特徴は以下のとおりである:前記ステップ 4 はさらに以下のステップを含む:一次骨粉を完成品骨粉に製造する:遠心噴霧した一次骨粉をさらに遊星ボールミル粉砕して粒径が 10 から 12μm の超微細な完成品骨粉にし,ここで,遊星ボールミルの回転速度は以下のとおりである:450 から 500r/min であり,研磨時間は 3 から 4h である;その後,内層は清潔なプラスチック袋であり,1 袋あたり 1Kg/袋で包装して骨粉製品を得て,密閉して保管し,乾燥箇所で日陰で保存する。
校正結果	7.請求項 1 に記載の勾配微粉砕同期熟成骨粉及び骨ペーストの加工方法であって,その特徴は,前記ステップ四がさらに以下のステップを含むことである:一次骨粉を完成品骨粉に調製する:遠心噴霧した一次骨粉をさらに遊星ボールミル粉砕して粒径が 10~12μm の超微細な完成品骨粉にし,ここで,遊星ボールミルの回転速度は以下のとおりである:450~500r/min であり,粉砕時間は 3~4h である;その後,内層に清潔なプラスチック袋を用い,一袋ごとに 1Kg/袋で包装して骨粉製品を得て,密封して入庫し,乾燥した日陰で保存する。
解　　説	中国語「研磨」は、日本語の「研磨」と「粉砕（挽き砕く）」の双方の意味があるため、技術内容・文脈に応じて訳し分ける必要がある。 本例は骨粉の加工に関し、原文中に二度出現する「研磨」はどちらも「粉砕」を意味する。しかし、機械和訳では、一方は「粉砕」と正しく訳せたものの、もう一方は「研磨」と誤訳されている。 ニューラル機械翻訳では、学習データ中に有力な訳語候補が複数存在する場合、このように、同一文中でも異なる訳語が採用されることがある。正しく訳せた方の「研磨」は、直前に「球磨机（ボールミル）」という粉砕器具が存在し、「粉砕」が採用されやすい状況であったのに対し、「研磨」と誤訳された「研磨」は、「研磨时间」という、粉砕と研磨のどちらにも該当し得る使われ方である。「研磨」が採用されたのは、機械翻訳プラットフォームの学習データ中では、「研磨」により近似のベクトル値が付されていたためと見られる。

［事例 3.2.4-4：CN107019420B ｾｸｼｮﾝ A］

原　　文	所述限压阀(50)和所述锅盖(20)之间设置有螺纹举升机构(70)，所述电机(81)驱动所述限压阀(50)旋转时，所述限压阀(50)在所述螺纹举升机构(70)的作用下上升或者下降，从而开启或者封闭所述排气通道(30)，
機械和訳	前記圧力制限弁(50)と前記鍋蓋(20)との間にねじ紋挙昇降機構(70)が設置され,前記モータ(81)が前記圧力制限弁(50)を回転駆動すると,前記圧力制限弁(50)は前記ねじ紋挙昇降機構(70)の作用で上昇又は下降し,それにより前記排気通路(30)を開閉し
校正結果	前記圧力制限弁(50)と前記鍋蓋(20)との間にねじリフト機構(70)が設けられ,前記モータ(81)が前記圧力制限弁(50)の回転を駆動する時,前記圧力制限弁(50)は前記ねじリフト機構(70)の作用で上昇又は下降し,それにより前記排気通路(30)を開閉し,
解　　説	翻訳文中に中国語「紋挙」がそのまま出力されている。旧方式の機械翻訳では、翻訳不能語は「未知語」としてしばしば原語（ここでは中国語）のまま出力されていた。ニューラル機械翻訳で用いられるトークンは、従来の機械翻訳が翻訳単位として用いていた形態素よりも細かく分割されるため、未知語の発生は大幅に抑制されるようになったが、稀に原語がそのまま出力されることもある。160 件のサンプルでも 13 案件で検出された。 「螺纹举升机构」は「螺纹」＋「举升」＋「机构」という複合語だが、機械和訳はこれを「螺（ねじ)」＋「纹举（紋挙)」＋「升（リフト)」＋「机构（機構)」と分解して翻訳したと推察される。「举」と「升」を分離させ、後者のみを「リフト」と対応づけるのは文法的に不正確だが、第 3.2.3 項のトークン分割実例からも分かるように、ニューラル機械翻訳のトークン分割では、こうした対応づけはよく見られる。

［事例 3.2.4-5：CN112832018B ｾｸｼｮﾝ D］

原　　文	所述的神经性化学战剂为沙林、塔崩、梭曼、维埃克斯或环沙林；
機械和訳	前記神経因性ケミカル兵器はサリン、タワー崩壊、ゼーマン、ベルル又はシクラムリンである;
校正結果	前記神経性化学兵器はサリン,タブン,ソマン,ブイエックス又はシクロサリンである;
解　　説	中国語では外来語の当て字を漢字で表記する。本例で列挙されている「沙林（サリン)」「塔崩（タブン)」「梭曼（ソマン)」「维埃克斯（ブイエックス)」「环沙林（シクロサリン)」はいずれも当て字に属する。 機械和訳は、「沙林」のみ「サリン」と正しく訳出されたが、他の四語

	は「タワー崩壊」「ゼーマン」「ベルル」「シクラムリン」と意味不明な誤訳となっている。 ニューラル機械翻訳では当て字も他の用語と同様、学習データ中に訳語が存在するか否かで訳出できるかどうかが決まる。各語の中日対応特許公報での使用例を調査したところ、「沙林」は 249 件と比較的多用されていたが、残る四語は 24 件、37 件、0 件、5 件と少数であり、学習データに十分な訳例が存在していない可能性が高い。 当て字は文字の意味ではなく発音に基づいている。このため、学習データ中に存在しない語を文字単位に分解して翻訳すると、本来の意味を全く類推できない重大な誤訳となる傾向が強い。 本例で誤訳された四語もこの傾向に該当する。まず「塔崩」→「タワー崩壊」は「塔」＋「崩」と過度に分解され、それぞれ文字の意味を踏まえて訳出されている。当て字では文字の意味は無関係なため、結果的に全く意味の異なる訳語となってしまっている。 続く三語は訳語の由来の推測は困難だが、「梭曼」→「ゼーマン」は、本来の「ゼーマン」の当て字「塞曼」と「曼」が共通するために採用された可能性がある。また、「维埃克斯」→「ベルル」も「ベルルカスト（维鲁司特)」と「维」が共通している。「环沙林」→「シクラムリン」に関しては、「环」は化合物の接頭辞「シクロ」として 30 万件以上の使用例があり（※当て字ではない)、かつ「沙林」は単独では正しく訳出されていたが、両者を組み合わせた「シクロサリン」とはならず、「シクラムリン」という意味不明な訳語となっている。「シクロ」は「三环唑（トリシクラゾール)」などで、「ムリン」も「橡胶环（ゴムリング)」などで、それぞれ「环」と対応づけられている可能性がある。 160 件のサンプルでは、当て字の誤訳に該当するケースが 2 件のみ検出された（C20、E16)。このうちサンプル E16 では「斯特封（ステップシール。ステップに相当する斯特が当て字)」と「格莱圈（グライドリング。格莱が当て字)」の 2 語が該当し、前者が「雌型シール」「メスシール」「スターシール」、後者が「グライダー」「レール」「レールリング」と、どちらも当て字部分が全く意味の異なる訳語に誤訳されている。

3.3. 訳抜け

訳抜けとは、原文の内容が翻訳文に反映されない不備を指す。主な原因として、原文中の用語や文言に対応する訳例が学習データ中に存在しないことや、入力文の不正なトークン分割により学習データ中の訳語が採用できなかったことなどが挙げられる。

3.3.1. 用語レベルの訳抜け

まず、用語レベルの訳抜けの事例を示す。

［事例 3.3.1-1：CN105574733B ｾｸｼｮﾝ G］

原　　文	当期销售器，用于根据初始采购量商品采购条款发布当期线下线上销售条款； 预售销售器，用于发布下一个销售周期线上商品不限量预售条款。
機械和訳	初期仕込み量商品調達条項に基づいてオンラインオフラインライン上販売条件を発行するために用いられる; 販売売り手は,次の販売周期ライン上の商品が無制限に販売される条項を発行する。
校正結果	現時販売器であって,初期購入量と商品購入条件に基づいて,現在のオフライン及びオンラインの販売条件を公開するために用いられる; 先行販売用販売器であって,次の販売サイクルのオンライン商品における無制限の先行販売条件を公開するために用いられる。

本例は、原文 1 行目の「当期销售器（現時販売器）」の箇所が機械和訳に訳出されていない。「当期销售器」は出願人の造語と考えられ、中日対応特許公報にも使用例はない。「当期」は一般語だが、「销售器」はマイナーな技術用語で、中日対応特許では 2 案件、6 文のみで使用されている（それらの訳語は「販売商品表示部」）。このため、機械翻訳プラットフォームの学習データ中に存在する確率は高くない。

ただし、本例では、一般語で豊富な訳例が存在するはずの「当期」も訳抜けしている。これは、ニューラル機械翻訳で各用語に付与されるベクトル値が、周辺の語句の状況を反映したものであることに起因すると考えられる。学習データ中に「当期」を含む訳例は多数存在しても、「销售器」と組み合わされた本例の状況に合致するベクトル値を持つものが存在せず、このためどれも採用されなかった可能性がある。もしくは、「销售器」という用語が学習データ中に存在しないため「当期」＋「销售器」と正しく分割できなかった可能性や、「当期」が修飾語であるため、修飾先の「销售器」が訳出されない限り単独では訳出できなかった可能性もある。

なお、「销售器」は 3 行目にも「预售销售器」として出現し、こちらは「販売売り手」と訳出されている。だが、実際の意味（「先行販売用販売器」）には程遠く、文字単位に分解したうえでその一部を訳出した結果と見られる（「销售」は本例でも各所で「販売」と訳されている）。

こうした難訳語の訳抜けは、サンプル 160 件では 9 件で検出された。サンプル D15 の「林蛙（アカガエル）」、E1 の「地勘（地質調査）」と「放线（墨出し）」、D11 の「洗可穿免烫整理（ウォッシュアンドウェアー仕上げ）」などは難訳語の典型例といえる。

一方、このような難訳語ではなく、本来ならば問題なく訳せるはずの用語が訳抜けするケースも存在する。一例を示す。

［事例 3.3.1-2：CN106459922B ｾｸｼｮﾝ C］

原　　文	1.一种心肌前体细胞，其是 CD82 为阳性，CD73 和 CD105 为阴性，且选自 ROR2、CD13、PDGFRα 和 KDR 中的至少一个细胞表面标记为阳性的细胞。
	：
	3.根据权利要求 1 所述的心肌前体细胞，其是选自下面的细胞表面标记 B 组中的至少一个细胞表面标记为阴性、和/或选自下面的细胞表面标记 C 组中的至少一个细胞表面标记为阳性的细胞，细胞表面标记 B 组：CD7、CD37、CD43、CD144、STRO-1、CD177、CD163，细胞表面标记 C 组：CD137L、CD140b、CD180、CD252、CD344、CD118、CD99R。
機械和訳	1.心筋前駆細胞であって,それは CD82 が陽性であり,CD73 及び CD105 が陰性であり,且つ ROR2,CD13,PDGFRα 及び KDR からなる群から選択される少なくとも一つのマーカーが陽性である細胞である。
	：
	3.請求項 1 に記載の心筋前駆細胞であって,それは以下の細胞表面マーカーB 群から選択される少なくとも一つの細胞表面マーカーが陰性であり,及び/又は以下の細胞表面マーカーC 群から選択される少なくとも一つの細胞表面マーカーが陽性である細胞であり細胞表面マーカーB 群:CD7,CD37,CD43,CD144,STRO－1,CD177,CD163 細胞表面マーカーC 群:CD137L,CD140b,CD180,CD252,CD344,CD118,CD99R。
校正結果	1.心筋前駆細胞であって,それは CD82 が陽性であり,CD73 及び CD105 が陰性であり,且つ ROR2,CD13,PDGFRα 及び KDR からなる群から選択される少なくとも一つの細胞表面マーカーが陽性の細胞である。
	：

	3.請求項 1 に記載の心筋前駆細胞であって,それは以下の細胞表面マーカーB 群から選択される少なくとも一つの細胞表面マーカーが陰性であり,及び/又は以下の細胞表面マーカーC 群から選択される少なくとも一つの細胞表面マーカーが陽性の細胞であり,細胞表面マーカーB 群:CD7,CD37,CD43,CD144,STRO－1,CD177,CD163,細胞表面マーカーC 群:CD137L,CD140b,CD180,CD252,CD344,CD118,CD99R。

本例では、「細胞表面マーカー」と訳すべき「細胞表面标记」が、請求項 1 では「マーカー」とのみ訳され、「细胞表面」が訳抜けしている。だが、後続の請求項 3 では同じ語が 6 回出現し、全て正しく訳されている。両請求項とも「细胞表面标记」の周辺の語句はほぼ同じで、請求項 1 が特殊な使用例とも言えない。本例から、本来正しく訳出できる用語が普通の使われ方をしていても、文の構造や用語の位置、共起する語のわずかな差異等によって訳抜けが生じ得ることがわかる。サンプル 160 件においても、こうした突発的な訳抜けは 47 件で検出され、難訳語の訳抜け（9 件）よりも高頻度で発生している。その大半は上記事例と同様、別の箇所では問題なく訳されている一般語である。サンプル B10 の「化合物」、C5 の「腫瘍」と「増加」、F15 の「駆動する」等が典型例として挙げられる。

本項冒頭で述べた訳抜けの主な原因、すなわち「訳例が学習データ中に存在しないこと」や「入力文の状況によるトークン化の失敗」は、前項で分析した誤訳の発生原因と共通する。実際、事例 3.3.1-1 の「销售器」や前項の事例 3.2.1-1 の「刷单行为」は、同じ用語が文によって誤訳されたり訳抜けしたりしていた。つまり、用語レベルの誤訳と訳抜けは、現象としては別種だが、どちらも翻訳不能な文言の存在によって生じる不備のバリエーションであり、根本的には同一の課題といえる。よって、その解決策も誤訳に準ずる。

3.3.2. 入力文字数超過による訳抜け

ただし、訳抜けには、誤訳とは異なる原因で発生するタイプも存在する。入力文の文長が機械翻訳エンジンの許容文字数／トークン数を超過したことによる訳抜けである。

第 3.1 項で述べたとおり、エンジンが 1 回の処理で取り扱えるトークン数には上限がある。機械翻訳プラットフォームの翻訳文の中には、入力文がこの上限を超過したことで大規模な欠落が生じたと見られるケースが存在する。一例を示す。

［事例 3.3.2-1：CN107846845B ｾｸｼｮﾝ C］

原　　文	7.根据权利要求 1 至 4 中任一项所述的捆包机，其中所述捆包机(1)设置有运动装置(99、85、87、185、187)，所述运动装置用于使为门的形式的所述进料器管道(23)的底壁(37)的第一门(59)和也为门的形式的所述进料器

	管道(23)的底壁(37)的第二门(61)选择性地在关闭位置和部分打开位置之间运动，在所述关闭位置中，所述第一门(59)的与其附接到所述输送通道的底壁(51)的边缘相对的边缘(67)与所述进料器管道(23)的底壁(37)的第二门(61)的边缘(69)相邻，由此与所述进料器管道(23)的底壁(37)的第二门(61)一起提供所述下边界的至少一部分，在所述部分打开位置中，所述进料器管道(23)的底壁(37)的第一门(59)和底壁(37)的第二门(61)远离所述进料器管道(23)的项壁(35)定位，由此使得所述下边界保持完整。
機械和訳	7.請求項 1～4 のいずれか一項に記載の梱包装置は,梱包装置(1)運動装置(99,85,87,185,187)が設けられており,前記運動装置使为门形態用の前記供給管(23)の底壁(37)の第一の扉(59)とも,ドアの形の前記供給パイプ(23)の底壁(37)の第二のドア(61)の閉位置と開位置との間で選択的に移動可能であり,前記閉鎖位置において,前記第一のゲート(59)の前記供給通路の底壁(51)に取り付けられている端部とは反対側の縁部(67)は,供給導管(23)の底(37)の第二(61)の縁部(69)に隣接して配置され,これにより,前記供給パイプ(23)の底壁(37)の第二ゲート(61)と前記下側境界の少なくとも一部を提供する,部分的に開いた位置では,前記供給管(23)の前記底壁(37)の第一

本例の機械和訳は文の途中で明らかに途絶している。具体的には、原文末尾部分（「门(59)和底壁(37)的…保持完整。」）に相当する訳文が出力されていない。同様の翻訳文の途絶は、CN106804542B の請求項 1、CN106879564B の請求項 1、CN106982860B の請求項 1 など、本調査対象中の少なくとも数十例で見られた[48]。

いずれも原文の途中から末尾までが完全に訳抜けしており、原文を読み込む際に文字数が上限を超え、それ以降が読み込めなかった状況が窺える。上例の原文は 317 文字までで途切れているが、上掲の 3 案件もそれぞれ 307 文字、320 文字、347 文字と、おおむね同程度の文字数で途絶している[49]。

類似の事例をもう一例示す。

[48] 機械和訳の末尾が句点「。」でないという条件に該当するものから目視で検出した。

[49] ニューラル機械翻訳では、1 文として扱えるトークン数に上限がある。これらの事例もトークン数の上限に抵触している可能性が高いが、入力文がどのようにトークン分割されているかは外部からは知り得ない。

［事例 3.3.2-2：CN107505041B ｾｸｼｮﾝ G］

原　　文	1.基于相位敏感光时域反射计的相位解调方法，其特征在于，该方法基于相位解调装置，所述相位解调装置包括窄线宽激光器、第一光耦合器、光调制器、光放大器，环形器、第二光耦合器、传感光纤、脉冲信号发生器、光电探测器和数据采集器，所述传感光纤中均匀间隔刻有弱反射光栅，第二光耦合器为 50:50 光耦合器，脉冲信号发生器用于控制光调制器和数据采集器；窄线宽激光器发出连续光，该连续光经第一耦合器分成功率不等的两路光，其中功率较小的一路光作为本地参考光，而功率较大的一路光作为探测光，该探测光经光调制器调制成探测双脉冲，该探测双脉冲经光放大器放大后输入光环行器的第一端口，并由光环行器的第二端口注入传感光纤，探测双脉冲在传感光纤中产生的背向返回信号经光环行器的第三端口输入第二光耦合器，第二光耦合器将背向返回信号与本地参考光干涉，得到与光调制器移频数值相等的拍频信号，光电探测器将拍频信号转换为电信号，数据采集器对该电信号进行采集和处理，获得传感光纤的扰动信息；
機械和訳	1.位相敏感光時間領域反射計に基づく位相復調方法であって,その特徴は以下のとおりである:該方法は位相復調装置に基づき,前記位相復調装置が狭い線幅レーザ、第一光カプラ、光変調器、光増幅器を含み,サーキュレータ、第 2 光カプラ、センシングファイバ、パルス信号発生器、光検出器及びデータ収集器を含み,前記センシングファイバには弱い反射格子が均一に間隔をおいて刻まれ,第 2 光カプラは 50:50 光カプラであり,パルス信号発生器は光変調器及びデータ収集器を制御するために用いられる;狭線幅レーザは連続光を発し,該連続光は第一カプラによってパワーが異なる 2 パス光に分割され,ここでパワーが小さい方の光はローカル基準光として,パワーが大きい方の光はプローブ光として,該検出された電気信号を収集及び処理するために,電気信号に基づいて以下のステップを行う:データ収集器が収集した電気信号に基づいて以下のステップを行う:光変調器の周波数変換値と等しいビート周波数信号を取得するために,電力が大きい方の光をプローブ光として,該検出は光変調器が収集した電気信号に基づいて以下のステップを行う:光変調器により収集された電気信号に基づいて以下のステップを行う:電力がより大きい光を検出光とし,該検出光を光変調器により収集した電気信号に基づいて取得する。2 3 2 2。
校正結果	1.位相敏感光時間領域反射計に基づく位相復調方法であって,その特徴は以下のとおりである:該方法は位相復調装置に基づき,前記位相復調装置は狭線幅レーザ,第一光カプラ,光変調器,光増幅器,サーキュレータ,第二光カプラ,センシングファイバ,パルス信号発生器,光検出器及びデータ収集器

	を含み,前記センシングファイバに弱反射格子が均一に間隔を置いて刻まれ,第二光カプラは 50:50 光カプラであり,パルス信号発生器は光変調器及びデータ収集器を制御するために用いられる;狭線幅レーザは連続光を発し,該連続光は第一カプラによってパワーが異なる二パス光に分割され,ここでパワーが小さい方の光はローカル参照光とし,パワーが大きい方の光はプローブ光とし,該検出光は光変調器によって検出ダブルパルスに変調され,該検出ダブルパルスは光増幅器で増幅された後に光サーキュレータの第一ポートに入力され,且つ光サーキュレータの第二ポートからセンシングファイバに注入され,検出ダブルパルスによってセンシングファイバで生成されたバックリターン信号は光サーキュレータの第三ポートを経て第二光カプラに入力され,第二光カプラはバックリターン信号をローカル参照光と干渉させ,光変調器の周波数シフト値に等しいビート周波数信号を取得し,光検出器はビート周波数信号を電気信号に変換し,データ収集器は該電気信号に対して収集及び処理を行い,センシングファイバの摂動情報を取得する;

本例は、セミコロンで 6 分割された請求項の 1 文目で、423 文字と長大である。機械和訳の文末は「電気信号に基づいて取得する。」と句点で結ばれており、前例のような途絶は発生していないように見える。だが、実際には、機械和訳の後半（赤字部分）は、原文の後半（青字部分）とは全く異なる内容になっている。赤字部分冒頭の「該検出された電気信号を収集及び処理するために,電気信号に基づいて以下のステップを行う:」は原文末尾付近の「数据采集器对该电信号进行采集和处理（データ収集器は該電気信号に対して収集及び処理を行い)」と文言の一部が共通し、「光変調器の周波数変換値と等しいビート周波数信号を取得するために」も原文の「得到与光调制器移频数值相等的拍频信号,（光変調器の周波数シフト値に等しいビート周波数信号を取得し,)」に相当するが、それ以外の原文の内容は訳出されておらず、既出の内容が繰り返されているのみである。つまり、見かけ上は完結しているが、実際には文の後半の大部分が訳抜けしており、前例（文の後半で途絶）と同様の状況といえる。サンプル 160 件でも 11 件で検出され、その大半は上記事例 3.3.2-2 と同様の、ひとまず翻訳文は完結しているが実際には大規模な訳抜けが生じているタイプ（F12 の請求項 1 など）であったが、文末が唐突に途切れる事例 3.3.2-1 と同タイプのものも 1 件存在した（D9 の請求項 7)。

これら二例の共通点として、文がセミコロンや意図的改行で区切られず長大になっている（またはセミコロンで区切られた後も長大である）ことが挙げられる。請求項の中には文長がさらに長いものも多数存在するが、大抵は構成要素やステップなど適度な単位でセミコロンや意図的改行によって区切られている。機械翻訳プラットフォームの機械翻訳エ

ンジンは、こうしたセミコロンや意図的改行で請求項を適宜分割して個別に翻訳していると考えられる。このため、請求項自体は長大でも、翻訳処理単位という観点からは、適度な文長の文の集合体と見なせるものが多い。

これに対し、翻訳文が途絶した上掲の二例は、いずれもセミコロンや意図的改行が用いられておらず、エンジンが分割できる区切りが存在しない[50]。このため翻訳処理単位が長大となり、機械翻訳エンジンが扱えるトークン数の上限に達して、以降の翻訳文が生成できなくなったと推察される。

3.3.3. セミコロン区切り文での訳抜け

ただし、請求項がセミコロン等で短く区切られていても、大きな訳抜けが発生するケースもある。以下、一例を示す。

［事例 3.3.3-1：CN106620893B ｾｸｼｮﾝ A］

原　　文	① 1.人工晶状体，其包括：
	② - 基体材料，选自疏水型丙烯酸酯、丙烯酸酯水凝胶、硅胶、硅水凝胶、氟硅丙烯酸酯、聚苯乙烯和聚甲基丙烯酸甲酯、聚碳酸酯、聚硅氧烷，或其混合物；
	③ - 至少一种光敏剂，选自光动力型和/或光热型光敏剂，光敏剂含有可聚合基团；
	④ 其中，光敏剂与基体材料以共价键方式结合，选自：
	⑤ - 光敏剂在基体材料成型过程中参与聚合；和/或
	⑥ - 光敏剂以表面接枝、表面修饰方式固定在基体材料表面，
	⑦其中基于材料总重量，所含光敏剂含量为小于 1 重量%。
機械和訳	① 1.眼内レンズであって,それは以下を含む:
	② -疎水性アクリレート,アクリレートヒドロゲル,シリカゲル,シリコーンヒドロゲル,フルオロシリコーンアクリレート,ポリスチレンおよびポリメチルメタクリレート,ポリカーボネート,ポリシロキサンからなる群から選択されるマトリックス材料;
	③ または光力学および/または光熱型光増感剤からなる群から選択される少なくとも 1 種の光増感剤;
	④ 光活性剤および/または光活性剤からなる群から選択される少なくとも 1 種の光増感剤;

[50] 事例 3.3.2-2 は請求項自体はセミコロンで 6 分割されているが、分割後の 1 文目が 423 文字と長大である。

	⑤ および/または
	⑥ -光増感剤が表面グラフト化され,表面修飾されて基材表面に固定される。
	⑦ この場合,含有される光増感剤の含有量は,材料の全質量に対して 1 質量%未満である。
校正結果	① 1.眼内レンズであって,それは以下を含む:
	② -マトリックス材料であって,疎水性アクリレート,アクリレートヒドロゲル,シリコーンゲル,シリコーンヒドロゲル,フルオロシリコーンアクリレート,ポリスチレンおよびポリメチルメタクリレート,ポリカーボネート,ポリシロキサン,またはその混合物から選択されるもの;
	③ -少なくとも 1 種の光増感剤であって,光力学型および/または光熱型から選択され,光増感剤が重合性基を含むもの;
	④ ここで,光増感剤はマトリックス材料と共役結合の方式で結合され,以下から選択される:
	⑤ -光増感剤がマトリックス材料の成型過程で重合に加わる;および/または
	⑥ -光増感剤が表面グラフト,表面修飾の方式でマトリックス材料表面に固定される;
	⑦ ここで材料の全重量に対して,含まれる光増感剤の含有量は 1 重量%未満である。

※考察の便宜上、各欄冒頭に①～⑦を付している。

本請求項はセミコロンと意図的改行によって 7 行に分割されている。機械和訳は、これら 7 行のうち④行目と⑤行目が訳抜けしている。④行目に相当する位置には「光活性剤および/または光活性剤からなる群から選択される少なくとも 1 種の光増感剤;」という訳文が出力されているが、これは③行目の内容が（不正確に）繰り返されたものであり、本来の④行目の内容「ここで,光増感剤はマトリックス材料と共役結合の方式で結合され,以下から選択される:」とは内容が大きく異なる。これを原文④行目が誤訳された結果と見るのは無理があり、④行目の訳抜けと、③行目の内容の重複出力（湧き出し）の併発と見なすのが妥当である[51]。⑤行目は、セミコロン後の「および／または」のみ出力されているが、「光敏剤在基体材料成型过程中参与聚合;（光増感剤がマトリックス材料の成型過程で重合に加わる;）」という本文全体が訳抜けしている。

[51] このように訳抜け部分を埋め合わせる形で文中の別の箇所が重複出力される事象については、第 3.4 項で詳述する。

本例の④行目、⑤行目は、文長は短く、特段の難訳語も存在しない。このため、翻訳難易度は低いと見られる。それにもかかわらず、セミコロン区切り全体にわたる大規模な訳抜けが発生している。機械翻訳プラットフォームの機械和訳では、このように意図的改行やセミコロンで適度な翻訳単位に区切られた請求項においても、翻訳単位（≒行）全体が欠落するレベルの大規模な訳抜けが発生するケースが多数見られた。サンプル 160 件においても、53 件でこのタイプの訳抜けが検出された。

その反面、本調査対象の全請求項 949,882 項で、請求項全体の訳抜けは発生していない。短文で書かれた請求項は一回の翻訳処理で全体が翻訳されているはずだが、その全体が訳抜けするというケースは見られなかった。この状況から、上例のような翻訳処理単位レベルの大規模な訳抜けは、複数に区切られて翻訳される長大な請求項で発生する事象と見なせる。第 2 章の統計分析でも、請求項が長文となるにつれ訳抜けの発生が顕著に増加しており、文長と訳抜けとの間には高い相関が見られる。

なぜこのような訳抜けが発生するかは確言できないが、訳抜けしている箇所の内容からは、語句の翻訳難易度の問題とは考えにくい。機械翻訳プラットフォームではセミコロン等で文を必ずしも分割していない可能性や、長大な請求項を複数に区切って翻訳処理し、その結果を連結して翻訳文を組み立てるという一連の処理の過程で、翻訳単位の一部が翻訳されなかったり、もしくは翻訳文が消失したりしている可能性が高い。

大きな訳抜けが発生した機械和訳によく見られる特徴として、原文にない句点（。）が文中に出力されることが挙げられる。通常、セミコロンで区切られた請求項は翻訳文も原文通りセミコロンで区切られた形で出力される。だが、本例は、⑥行目の末尾にセミコロンではなく原文に存在しない句点が出力されている[52]。

この句点は、エンジンが翻訳を一旦完了し、機械和訳を出力したことを示している。それが前処理での文分割結果に基づくものか、エンジンによる翻訳処理の過程で生じたものかは知り得ないが、本例では、⑥行目の翻訳を終えた時点で各行の翻訳結果が連結され、一文として出力されたと考えられる。

そして、このように請求項の途中で句点が出力された場合、そこまでに大きな不備が発生していても、それ以降の行は再び正しく翻訳されていることが多い。本例も、句点直後の⑦行目は正しく翻訳されている。これは、それまでの内容が一文として出力されたことで翻訳文の組み立て処理は一旦完了し、⑦行目は別の新たな文として翻訳されたためとも

[52] 前項で取り上げた事例 3.3.2-2 の機械和訳における文末の句点も、原文中には存在しない。

考えられる。

こうした結果から、本例に見られるようなセミコロン区切り単位の大規模な訳抜けの根本的な原因は、入力文が長大であることによるエンジンの翻訳処理もしくはその前後処理の失敗である可能性が高い。ただし、その根本原因は必ずしも入力文の物理的な文長にあるとは限らない。事実、本例よりも長大な請求項でも、訳抜けせず、途中に句点が出力されることもなく全体が訳出されているものは多い[53]。よって、こうした大規模な訳抜けが発生した請求項では、通常は適切に行われる文の分割処理が何らかの理由で失敗している、又は翻訳処理が何らかの理由で中断されるなどの不具合が生じていると推察される。その原因を究明し、改善することで、大規模な訳抜けの抑制が期待できる。この点に関しては第 3.7.4 項で再度考察する。

3.3.4. 事例集

本調査分析で検出した、訳抜けに該当する特徴的な事例を掲載する。

［事例 3.3.4-1：CN106164026B ｾｸｼｮﾝ C］

原　　文	24.用于制备 1,1,1-三氟-2,3-二氯丙烷的方法，其包括在以催化有效量的催化剂存在下使氯气与 3,3,3-三氟丙烯接触以形成 1,1,1-三氟-2,3-二氯丙烷，其中所述催化剂包含至少一种金属卤化物，其中所述金属是来自周期表中的第 13、14 或 15 族的金属或过渡金属或其组合，所述氯化反应在液相中进行，其中所述催化剂有效量为存在的反应物的总量的 0.1 至 10 重量%，其中所述金属为镍、铬、铁、钪、钇、镧、钛、锆、铪、钒、钼、钨、锰、铼、钌、锇、钴、钯、铜、锌、钽、铝、锡或铅。
機械和訳	24.触媒有効量の触媒の存在下で塩素を 3,3,3-トリフルオロプロペンと接触させて 1,1,1-トリフルオロ-2,3-ジクロロプロパンを形成する工程を含み,前記触媒が少なくとも一つの金属ハロゲン化物を含み,前記金属が周期表の第 13 族,第 14 族または第 15 族の金属または遷移金属またはこれらの組合せであり,前記金属が,存在する反応物質の全量の 0.1～10 重量%であり,前記金属が,ニッケル,クロム,鉄,スカンジウム,アルミニウム,スズまたは鉛であり,前記触媒が,存在する反応物の総量の～重量%である,1,1,1-トリフルオロ-2,3-ジクロロプロパンの製造方法。
校正結果	24.1,1,1-トリフルオロ-2,3-ジクロロプロパンの製造方法であって,触媒作

[53] 上例（CN106620893B の請求項 1）の原文文長は 217 文字であるが、例えば CN106872930B の請求項 1（1,007 文字）や CN106212132B の請求項 1（770 文字）をはじめ、極めて長大であるにもかかわらず全文が破綻なく訳出され、文中に句点も出力もないものは多数存在する。

	用有効量の触媒の存在下で塩素ガスを 3,3,3-トリフルオロプロペンと接触させて 1,1,1-トリフルオロ-2,3-ジクロロプロパンを形成することを含み,ここで前記触媒は少なくとも 1 種の金属ハロゲン化物を含み,ここで前記金属は周期表の第 13 族,第 14 族または第 15 族からの金属または遷移金属またはそれらの組み合わせであり,前記塩素化反応は気相中で行い,ここで前記触媒の有効量は存在する反応物の総量の 0.1～10 重量%であり,ここで前記金属はニッケル,クロム,鉄,スカンジウム,イットリウム,ランタン,チタン,ジルコニウム,ハフニウム,バナジウム,モリブデン,タングステン,マンガン,レニウム,ルテニウム,オスミウム,コバルト,パラジウム,銅,亜鉛,タンタル,アルミニウム,スズ又は鉛である。
解　　説	機械和訳では、原文に列挙された 24 種の金属名称のうち冒頭の 4 つと末尾の 3 つのみが訳出され、それら以外は訳抜けしている。 ただし、同案件の請求項 1 では、同様に列挙された 24 種の金属名称を正しく翻訳できていた。双方の相違点は、請求項 1 は意図的改行で区切られていて翻訳処理単位が比較的短い（77 文字）のに対し、本例には区切りがなく、翻訳処理単位がやや長め（244 文字）であることである。一文字の語が多数列挙されたためトークン数が過大となったか、もしくは構文が複雑になったことで処理に失敗し、訳抜けが発生した可能性が高い。

3.4. 湧き出し

湧き出しとは、原文中に存在しない内容・文言が機械和訳中に出力される不備を指す。

3.4.1. 一般的な湧き出し

湧き出しの典型的な発生パターンとして、原文の一部が訳抜けし、これを埋め合わせるように文中の別の部分が重複して出力されるケースが多数見られた。以下、一例を示す。

［事例 3.4.1-1：CN105197014B ｾｸｼｮﾝ B］

原　　文	19.根据权利要求 16 所述的方法，其中，确定步骤包括：通过融合器基于从行驶车道匹配器和行驶车道跟踪器输出的结果及结果模式确定所述行驶车道匹配结果或者所述行驶车道跟踪结果为所述车辆的所述当前行驶车道。
機械和訳	19.請求項 16 に記載の方法であって,ここで,決定ステップは以下を含む:前記走行車線マッチング結果又は前記走行車線追跡結果は,前記走行車線マッチング装置及び走行車線追跡装置から出力された結果及び結果パターンに基づいて前記走行車線マッチング結果又は前記走行車線追跡結果が前記車両の現在走行車線であるか否かを判断するためのものである。
校正結果	19.請求項 16 に記載の方法であって,ここで,判断ステップは以下を含む:融合部を介して,走行車線マッチング部及び走行車線追跡部から出力された結果及び結果パターンに基づいて,前記走行車線のマッチング結果又は前記走行車線の追跡結果が前記車両の現在の走行車線であると判断する。

本例では、原文の「通过融合器（融合部を介して）」に相当する訳語が存在せず、その位置には「前記走行車線マッチング結果又は前記走行車線追跡結果は」という全く異なる内容が出力されている。この内容は、原文後半の「所述行驶车道匹配结果或者所述行驶车道跟踪结果」に相当するが、この部分は本来の位置に「前記走行車線マッチング結果又は前記走行車線追跡結果が」と訳出されている。つまり、本例は「通过融合器」が訳抜けし、これを埋め合わせる形で原文の別の箇所「所述行驶车道匹配结果或者所述行驶车道跟踪结果」に相当する訳語が重複して湧き出している。同様の現象は前項までに示した事例 3.2.4-1、3.3.2-2、3.3.3-1 にも見られた。

このように、機械翻訳プラットフォームの機械和訳には、訳抜けが生じた箇所に別の箇所の翻訳結果が重複出力され、あたかも欠落を埋め合わせているかのように映るものが散見される。同様の事例をもう一つ示す。

［事例 3.4.1-2：CN112127093B ｾｸｼｮﾝ D］

原　　文	9.根据权利要求 7 或 8 所述的 X 射线图像采集系统(100)，其中，所述控制单元被配置为基于所述准直器配置来探测所述探测器表面(18)上的准直场(52)的几何结构；
	其中，所述几何结构包括初级延伸(54)和次级延伸(56)，所述次级延伸横向于所述初级延伸；
	并且所述初级延伸大于所述次级延伸；
	其中，所述探测器表面(18)具有一形状，所述探测器表面的所述形状具有初级探测器延伸和次级探测器延伸，并且所述初级探测器延伸大于所述次级探测器延伸；
	并且其中，所述旋转单元被配置为基于所述旋转信号(22)来旋转所述 X 射线探测器(16)，使得所述准直场(52)被限制约束到所述探测器表面(18)的所述形状。
機械和訳	9.請求項 7 又は 8 に記載の X 線画像収集システム(100)であって,ここで,前記制御ユニットは前記コリメータ構成に基づいて前記検出器表面(18)上のコリメートフィールド(52)の幾何学的構造を検出するように配置される;
	ここで,前記幾何学的構造は一次延長部(54)及び二次延長部(56)を含み,前記二次延長部は前記一次延長部より大きい;
	且つ前記一次延長部より大きい; 前記一次延長部は前記二次延長部より大きい;
	ここで,前記検出器表面(18)は形状を有し,前記検出器表面の前記形状は一次検出器延伸及び二次検出器延伸を有し,且つ前記一次検出器延伸は前記二次検出器延伸より大きい;
	またこの場合,回転ユニットは,回転信号(22)に基づいて X 線検出器(16)を回転させるように構成されており,その結果,コリメート磁場(52)が検出器面(18)の形状に拘束される。
校正結果	9.請求項 7 又は 8 に記載の X 線画像収集システム(100)であって,ここで,前記制御ユニットは前記コリメータ構成に基づいて前記検出器表面(18)のコリメーションフィールド(52)の幾何学的構造を検出するように構成される;
	ここで,前記幾何学的構造は一次延伸部(54)及び二次延伸部(56)を含み,前記二次延伸部は前記一次延伸部に対して横向きである;
	且つ前記一次延伸部は前記二次延伸部より大きい;
	ここで,前記検出器表面(18)は形状を有し,前記検出器表面の前記形状は一次検出器延伸部及び二次検出器延伸部を有し,且つ前記一次検出器延伸部

	は前記二次検出器延伸部より大きい;
	且つここで,前記回転ユニットは前記回転信号(22)に基づいて前記 X 線検出器(16)を回転させることで,前記コリメーションフィールド(52)を前記検出器表面(18)の前記形状に制限するように構成される。

本例は請求項がセミコロンで 5 文に区切られている。不備が発生しているのは 2 文目と 3 文目である。

まず、原文 2 文目の「所述次级延伸横向于所述初级延伸（前記二次延伸部は前記一次延伸部に対して横向きである）」に相当する部分が訳抜けし、代わりに「前記二次延長部は前記一次延長部より大きい」という、異なる内容が出力されている。この内容は、機械和訳の 3 文目とほぼ同じであり（ただし「一次」と「二次」が入れ替わっている）、訳抜け部分の埋め合わせとして別の箇所の内容が繰り返されるパターンに合致する。

なお、機械和訳 3 文目の冒頭に、2 文目で湧き出した文言の一部「前記一次延長部より大きい」が再度出力されている。セミコロンの区切りを超えて湧き出しが生じていることから、2 文目と 3 文目は同じ翻訳単位で処理されたと考えられる。このことから、機械翻訳プラットフォームでは、セミコロンでの文分割は必要が生じた場合のみに行われていることがわかる。

このような訳抜けとその埋め合わせ的な湧き出しの併発は、サンプル 160 件中 39 件で検出されており、（訳抜けを伴わない）単発の湧き出し 22 件よりもむしろ多数であった。この結果からも、訳抜けと湧き出しの併発は機械翻訳プラットフォームの典型的な不備パターンといえる。

訳抜けと湧き出しの併発は、必要な情報が失われ、かつ余分な情報が与えられるため、請求項の内容理解に重大な悪影響を及ぼす。根本的な発生原因は文の一部が訳抜けしたことにあり、それが湧き出しを誘発していると考えられる。長大な請求項で発生し易い不備であることは確かだが（第 2 章第 2.5 項参照）、必ずしも難訳語を含む文や複雑な構造の文で発生するとは限らず、発生条件の特定は難しい。

3.4.2. 不要な特許用語の湧き出し

請求項の機械和訳に特有の湧き出しのパターンとして、頻出の特許用語・表現が文中の用語や文脈から連想される形で湧き出すケースが散見された。以下に例示する。

［事例 3.4.2-1：CN107404922B ｾｸｼｮﾝ A］

原　　文	17.包含通过根据前述权利要求中任一项所述的方法获得的乳基蛋白质水解物的组合物。
機械和訳	17.請求項 1～10 のいずれか一項に記載の方法により得られた乳ベースタンパク加水分解物を含む組成物。
校正結果	17.前記請求項のいずれか一項に記載の方法により得られた乳ベースタンパク加水分解物を含む組成物。

［事例 3.4.2-2：CN106995964B ｾｸｼｮﾝ D］

原　　文	15.根据权利要求 8 所述的开口机器(2；1002)，其特征在于，所述间隔件具有上部和下部，并且在上部和下部处沿所述纵向轴线(X10，X1010)成对地重叠。
機械和訳	15.請求項 8 に記載の開口機器(2;1002)前記スペーサは上部および下部を有し、前記長手軸に沿って(X10，X1010)対をなして上下に重なり合うように上部および下部であることを特徴とする、請求項 1 に記載の方法。
校正結果	15.請求項 8 に記載の開口機械(2;1002)であって,その特徴は以下のとおりである:前記スペーサは上部および下部を有し,且つ上部と下部は前記縦方向軸線(X10,X1010)に沿って対をなすように重なり合う。

事例 3.4.2-1 では、原文が「前述权利要求中任一项所述（前記請求項のいずれか一項に記載の）」と、特定の請求項番号を記していないのに対し、機械和訳は「請求項 1～10 のいずれか一項に記載の」と、原文に書かれていない請求項番号が湧き出している。事例 3.4.2-2 も、機械和訳の末尾に、「請求項 1 に記載の方法」という、原文に存在せず、内容的にも無関係な文言が湧き出している。

こうした湧き出しは、突発的ながら一定の頻度で発生している。例えば事例 3.4.2-1 の「前述权利要求中任一项所述」という文言は本調査対象の請求項のうち 1,679 項に使われているが、ここから適訳語で訳されているもの（「前記請求項のいずれか一項に記載の」や、その同義表現）を除いた 129 件[54]に、同種の湧き出しが発生していた。事例 3.4.2-2 の

[54] CN107205415B の請求項 17、CN107920555B の請求項 7、CN106455859B の請求項 7、CN106618245B の請求項 11、CN105828772B の請求項 5 ほか多数。

「請求項 1 に記載の方法」の湧き出しに関しても、少なくとも原文に「权利要求 1」を含まず、かつ機械和訳が「請求項 1 に記載の方法。」で終わる請求項 73 件では同様の湧き出しが生じていると見なせる[55]。

これらのケースは、訳抜けがなく、かつ原文中に存在しない文言が湧き出していることから、訳抜けの埋め合わせに文中の内容を繰り返すタイプの湧き出しとは性質が異なる。湧き出す文言が請求項で多用されるものであることから、請求項由来の学習データの精度の問題である可能性がある。

機械翻訳プラットフォームの学習データには中日対応特許由来のコーパスが使用されている可能性が高い。しかしながら、対応特許であっても双方が必ずしも完璧な対訳文になっているとは限らない。特に特許請求の範囲は、請求項の数や順序が変わっていたり、記載内容や表現・様式が微妙に違ったりして、不正確な対訳文となりがちである。本項で挙げた二例は、湧き出している文言が請求項で多用されるものであることから、こうした不正確な対応関係の請求項に由来する学習データが原因である可能性がある。その場合、これら誤った対応関係の文例を除去することで改善が期待できる。

3.4.3. 事例集

本調査分析で検出された、湧き出しに該当する特徴的な不備の事例を掲載する。

［事例 3.4.3-1：CN107406889B ｾｸｼｮﾝ C］

原　　文	18.如权利要求 17 所述的方法，其中该文库被包含在嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、克劳氏芽孢杆菌、凝结芽孢杆菌、坚硬芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、短小芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌或苏云金芽孢杆菌宿主细胞内。
機械和訳	18.請求項 17 に記載の方法であって,その特徴は以下のとおりである:該ライブラリーはバチルス・セレウス、バチルス・アミロリケファシエンス、バチルス・セレウス、バチルス・サーキュラ、バチルス・クラウジ、バチルス・セレウス、バチルス・セレウス、バチルス・リケニフォルミス、バチルス・リケニフォルミス、バチルス・リケニフォルミス、バチルス・メガテリウム、バチルス・メガテリウム、バチルス・サーモフィルス、バチルスサチリス又はバチルス・チューリンゲンシス宿主細

[55] 原文に「权利要求 1」を含まず、かつ機械和訳に「請求項 1」を含む請求項 4,297 件の大半でも、類似の湧き出しが発生していると考えられる。

	胞内に含まれる。
校正結果	18.請求項 17 に記載の方法であって,ここで該ライブラリーはバチルス・アルカロフィラス,バチルス・アミロリケファシエンス,バチルス・ブレビス,バチルス・サーキュランス,バチルス・クラウジ,バチルス・コアグランス,バチルス・フィルムス,バチルス・ロータス,バチルス・レンタス,バチルス・リケニフォルミス,バチルス・メガテリウム,バチルス・プミルス,バチルス・ステアロサーモフィルス,バチルス・サブティリス又はバチルス・チューリンゲンシス宿主細胞内に含まれる。
解　　説	原文には「〇〇芽孢杆菌」という菌名が 15 種類列挙されている。いずれも専門性が高く、機械和訳ではその大半が誤訳されたが、異なる菌名に対して「バチルス・セレウス」が 4 回（正解なし)、「バチルス・リケニフォルミス」が 3 回（うち正解 1 回)、「バチルス・メガテリウム」が 2 回（うち正解 1 回)、繰り返し採用されている。単なる誤訳ではなく、翻訳不能な箇所を文中の他の語句で埋め合わせるタイプの湧き出し（⇒第 3.4.1 項）である可能性が高い。

［事例 3.4.3-2：CN112122652B ｾｸｼｮﾝ B］

原　　文	4.根据权利要求 3 所述的掏孔器，其特征在于，驱动件包括拉簧(11)和竖直滑动设置在凹槽(1b)内的驱动杆(12)，拨片(8)位于驱动杆(12)和限位杆(7)之间，拉簧(11)两端分别连接拨片(8)和桶体(1)，驱动杆(12)上设有从上到下依次分布的抵压面一(12b)和限位面一(12a)，限位杆(7)上设有从上到下依次设置的限位面二(7a)和抵压面二(7b)，在拉簧(11)作用下，拨片(8)上端的上下两侧分别抵压在抵压面一(12b)和限位面一(12a)上，拨片(8)下端的上下两侧分别抵压在限位面二(7a)和抵压面二(7b)上，驱动杆(12)上成型有水平设置的压杆(12c)以及设置有使驱动杆(12)具有向下运动趋势的第三弹簧(13)，且压杆(12c)外端伸出凹槽(1b)。
機械和訳	4.駆動ロッド(12)と制限レバー(7)との間に位置し,引張バネ(11)2 の両端がそれぞれタブ(8)とバレル(1)とに接続され,駆動ロッド(12)上に水平に配置された加圧ロッド(シンノリン c)及び駆動ロッド(タードン)上に水平に設置された加圧ロッド(シンキン c)と,駆動ロッド(タードン)上に水平に配置された加圧ロッド(シンマ c)と,駆動ロッド(タードン)上に水平に設置された加圧ロッド(シンマ c)と,駆動ロッド(タードン)上に水平に設置された加圧ロッド(シンキン c)と,駆動ロッド(タードン)上に水平に設置された加圧ロッド(シンキン c)と,駆動ロッド(タードン)上に水平に設置された加圧ロッド(シンキン c)と,駆動ロッド(タードン)上に水平に配置された加圧ロッド(シンキン c)と,駆動ロッド(タードン)上に形成され

	て水平に設置された加圧ロッド 12 12 7 2 7 2 3 7 11 8 2 12 12 8 2 2 7 11 2 7 12 12 12 3 13 12 1 1 1
校正結果	4.請求項 3 に記載のボアビットであって,その特徴は以下のとおりである:駆動ロッドは引張バネ(11)と凹部(1b)内に垂直に摺動するように設置された駆動ロッド(12)を含み,パドル(8)は駆動ロッド(12)と位置決めロッド(7)との間に位置し,引張バネ(11)の両端はそれぞれパドル(8)とバレル(1)とに接続され,駆動ロッド(12)上には上から下に順次押圧面一(12b)と位置決め面一(12a)が設けられ,位置決めロッド(7)に上から下に順次位置決め面二(7a)と押圧面(7b)が設置され,引張バネ(11)の作用下で,パドル(8)の上端の上下両側はそれぞれ押圧面一(12b)と位置決め面一(12a)に押付けられ,パドル(8)の下端の上下両側はそれぞれ位置決め面二(7a)と押圧面(7b)に押付けられ,駆動ロッド(12)には水平に設置された押圧ロッド(12c)が成形され駆動ロッド(12)に下向きに付勢する第三バネ(13)を有し,且つ押圧ロッド(12c)の外端は凹溝(1b)から突き出る。
解　　説	本例の機械和訳は、原文冒頭部以外の内容がほとんど訳出されておらず、唯一、原文後半の「驱动杆(12)上成型有水平设置的压杆(12c)」に相当する「駆動ロッド上に水平に配置された加圧ロッド」という文言が何回も繰り返されており、典型的な「訳抜けと、その埋め合わせとしての湧き出し」に属する。 ただし本例は、繰り返し部分において要素付番であるべき箇所に「シンノリン」「タードン」「シンキン」「シンマ」という意味不明の語が湧き出している。調査したところ、「シンノリン」は化学物質名（噌啉）として中日対応特許でも多数使用されていたが、他の三語は使用例がなかった。 他の案件で機械和訳に「タードン」「シンキン」「シンマ」を含むものが存在するかを調べた結果、要素付番に置き換わって「タードン」が湧き出しているケースが 3 件（CN111636399B の請求項 1 など)、「シンマ」が湧き出しているケースが 4 件（CN108720168B の請求項 1 など）検出された。これらの案件では他の要素付番も「¥」「ツボ」「ピアノ」という無関係な語に置き換わっているものが多かった。 上記調査結果のとおり、要素付番箇所に無関係な語句が湧き出す本例のような事象はごく限定的な発生にとどまっている。このため翻訳品質全体への影響は小さいが、特殊な事例として報告しておく。 なお、本事例の機械和訳の文末には、訳抜け箇所の原文に存在した数字が羅列されている。この事象については「数値の処理」の不備として次項（第 3.5.1 項）で分析する。

3.5. 数値の処理

本項では、原文中の数値の処理に関連する不備を分析する。

3.5.1. 文末への数値出力

機械翻訳プラットフォームの翻訳文を見ると、訳抜けが発生しても、その部分に含まれる数字のみは文の末尾にまとめて出力されている。以下、一例を示す。

［事例 3.5.1-1：CN110723907B ｾｸｼｮﾝ C］

原　　文	7.如权利要求 5 或 6 所述的方法，其特征在于，步骤 1)中，以 8℃/分钟的速率升温至 500℃，恒温 1 小时，以 4℃/分钟的速率继续升温至 1300℃，并恒温 1 小时，使原料熔融成液态，将坩埚中的液体倾倒在已预热的模具上固化成型，并迅速放入 350℃马弗炉中退火，恒温 1 小时，随炉降至室温，制得铕掺杂透明玻璃。
機械和訳	7.請求項 5 又は 6 に記載の方法であって,その特徴は以下のとおりである:ステップ 1)において,8℃/分の速度で 500℃まで昇温し,1 時間恒温で,4℃/分の速度で 1300℃まで昇温し続け,且つ 1 時間恒温で,原料を液状に溶融させて,ユーロピウムドープ透明ガラスを製造する。350 1。
校正結果	7.請求項 5 又は 6 に記載の方法であって,その特徴は以下のとおりである:ステップ 1)において,8℃/分の速度で 500℃まで昇温し,1 時間恒温し,4℃/分の速度で引き続き 1300℃まで昇温し,且つ 1 時間恒温し,原料を液状に溶融させ,坩堝中の液体を予熱済みの型に流して硬化成型し,且つ速やかに 350℃のマッフル炉に入れて焼鈍し,1 時間恒温し,炉に伴って室温まで降温し,ユーロピウムドープ透明ガラスを得る。

本例の機械和訳では、原文の「将坩埚中的液体倾倒在已预热的模具上固化成型，并迅速放入 350℃马弗炉中退火，恒温 1 小时，随炉降至室温」という部分が訳抜けしているが、ここに含まれる「350℃」と「1 小时」の数字部分のみは文末に「350 1」と出力されている。類似例をもう一件示す。

［事例 3.5.1-2：CN107795872B ｾｸｼｮﾝ F］

原　　文	5.根据权利要求 4 所述的一种可伸缩型便携灯具，其特征在于活动机构（6）和灯泡座底(5)之间设有伞骨机构（4），所述伞骨机构（4）包括有连接在多个伞骨支柱一（54）上的短伞骨（41），在多个伞骨支柱二（64）上设有长伞骨（42），在长伞骨（42）下端设有伞骨连接点（43），所述短

	伞骨（41）一端连接伞骨连接点（43），在所述长伞骨（42）上设有多个 LED 小灯（44）。
機械和訳	5.請求項 4 に記載の伸縮式携帯照明器具であって,その特徴は以下のとおりである:可動機構(6)と電球ベース(5)との間に傘機構(4)が設けられ,前記傘機構(4)は複数の傘支柱(54)に接続される短傘(41)を含み,複数の傘支柱 2(64)に長傘(42)が設けられ,長傘(41)の端部には傘骨接続点(43)が設けられ,前記長傘(42)の一端には傘骨接続点()が設けられ,前記長傘骨()には複数の LED 小灯(44)が設けられる。42 43。
校正結果	5.請求項 4 に記載の伸縮可能な携帯照明器具であって,その特徴は以下のとおりである:可動機構(6)と電球ベース(5)との間に傘骨機構(4)が設けられ,前記傘骨機構(4)は複数の傘骨柱 1(54)に接続される短傘骨(41)を含み,複数の傘骨柱 2(64)に長傘骨(42)が設けられ,長傘骨(42)の下端に傘骨接続点(43)が設けられ,前記短傘骨(41)の一端に傘骨接続点(43)が設けられ,前記長傘骨(42)に複数の LED 小型ランプ(44)が設けられる。

本例の機械和訳では、「傘骨接続点()」と「前記長傘骨()」の要素付番がカッコ内にセットされず、文末に「42 43」と出力されている[56]。

こうした事例から、機械翻訳プラットフォームのエンジンには、完成させた翻訳文と原文を比較し、足りない数字を末尾に出力しているか、もしくは数字をプレースホルダー方式で翻訳文に反映させていると考えられる。

ニューラル機械翻訳方式では、原文中の数値を翻訳文に正確に転記する一手段として、原文中の数値を前処理でプレースホルダー（戻すための印）に置き換え、プレースホルダー化した原文を機械翻訳処理して、最後に、翻訳文中に組み込まれたプレースホルダーを数値に置き換え直すという一連の処理が用いられることがある。これにより、学習データの影響等による数値の不慮の変化を回避できる。

機械和訳の状況から、機械翻訳プラットフォームの機械翻訳エンジンもプレースホルダー方式を採用している可能性がある。上掲の二例も、プレースホルダーに置換されてエンジン内に記憶された原文中の数字が、訳抜け等によって翻訳文がプレースホルダーごと消

[56] 機械和訳を詳細に見ると、原文の「在长伞骨（42）下端设有伞骨连接点（43）（長傘骨(42)の下端に傘骨接続点(43)が設けられ)」と「所述短伞骨（41）一端连接伞骨连接点（43）（前記短傘骨(41)の一端に傘骨接続点(43)が設けられ)」の出力順序が逆になっている。さらに後半の主語「所述短伞骨（41）（前記短傘骨(41))」が「長傘(41)」となっているなど、出力内容が大きく乱れており、単なる要素付番の配置ミスではない。

失して行き場を失った結果、やむなく文末に羅列されたとも解釈できる。

こうした機械翻訳文末尾の数値の羅列は、数値を含む部分の訳抜けなど翻訳文生成の失敗が根本的原因であって、数値処理自体の問題ではない。基本的には、機械翻訳文が正しく生成されれば、原文中の数字も正確に反映されるといえる。ただし現状では、サンプル160 件中の 26 件で訳抜けに起因する文末の数値羅列が生じている。

3.5.2. 漢数字の処理に係る不備

その一方で、エンジンの数値処理の問題と考えられる不備も検出されている。それは、漢数字の処理に関する。以下、一例を示す。

［事例 3.5.2-1：CN109736014B ｾｸｼｮﾝ D］

原　　文	7.根据权利要求 6 所述的一种六边形 FRP 筋材网片的制造方法，其特征在于，所述限位块的材质为聚四氟乙烯、聚甲醛或硬质硅胶。
機械和訳	7.請求項 6 に記載の 6 角形 FRP 筋材メッシュの製造方法であって,その特徴は以下のとおりである:前記ストッパブロックの材質はポリテトラフルオロエチレン,ポリオキシメチレン又は硬質シリカゲルである。4。
校正結果	7.請求項 6 に記載の六角形 FRP リブメッシュの製造方法であって,その特徴は以下のとおりである:前記位置決めブロックの材質はポリテトラフルオロエチレン,ポリオキシメチレン又は硬質シリカゲルである。

本例の機械和訳は、文末に「4」と出力されている。原文中に「4」という算用数字は存在せず、「聚四氟乙烯」という化合物中に漢数字「四」が存在するのみである。同様のケースが多発していることから、機械翻訳プラットフォームには、文中の漢数字から算用数字を生成して記憶する処理が組み込まれていると考えられる。

この処理は、本例の別箇所の「六边形」→「6 角形」など、漢数字部分をそのまま算用数字に置換できるケースでは正しく機能する。だが、「聚四氟乙烯」のように漢数字が用語の一部として用いられ、算用数字に置換すべきでないケースでこの処理が働くと、生成された算用数字が行き場を失い、文末に出力されると推察される。この不備は機械翻訳プラットフォームでは頻発しており、サンプル 160 件でも 12 件で検出された。

なお、この現象は漢数字に準ずる漢字でも発生する。とりわけ「两（両）」という文字で顕著であり、例えば「两侧」が「両側」と訳出される一方で、別途派生した「2」が文

末に出力されるケースが多発している[57]。サンプル 160 件でも 17 件がこれに該当した。

また、漢数字から算用数字が派生することで、原文中の数字よりも多くの数字が存在することとなり、翻訳文で数値のずれが生じる場合がある。以下のようなケースである。

［事例 3.5.2-2：CN111826764B ｾｸｼｮﾝ D］

原　　文	S1.称取四氢呋喃与 N,N-二甲基甲酰胺混合至均匀，形成混合溶剂，加入所述改性聚丙烯腈，搅拌至均匀后，形成改性聚丙烯腈混合液；
機械和訳	S4.テトラヒドロフランと N,N-ジメチルホルムアミドとを均一に混合し,混合溶媒を形成し,前記変性ポリアクリロニトリルを加え,均一に撹拌した後,変性ポリアクリロニトリル混合物を形成する;
校正結果	S1.テトラヒドロフランと N,N-ジメチルホルムアミドとを均一に混合し,混合溶媒を形成し,前記改質ポリアクリロニトリルを添加し,均一に撹拌することで,改質ポリアクリロニトリル混合液を形成する;

本例では、原文のステップ番号「S1」が機械和訳では「S4」と誤っている。この現象は、機械翻訳プラットフォームがプレースホルダー方式を採用していると仮定すると説明がつき易い。つまり、原文中の「1」がプレースホルダーに置換されたが、これに加えて「四氢呋喃（テトラヒドロフラン）」中の漢数字「四」から算用数字 4 も派生し、これが「1」よりも優先して翻訳文中のプレースホルダーにセットされたとも解釈できる[58]。

さらに、漢数字から派生した算用数字が、原文と無関係な文言に組み込まれて翻訳文中に湧き出す事例も見られる。

［事例 3.5.2-3：CN109044881B ｾｸｼｮﾝ A］

原　　文	4)投入非硫酸盐两性表面活性剂、非硫酸盐非离子表面活性剂，搅拌均匀，降温至 40-35℃；出料。
機械和訳	4)は非硫酸塩性界面活性剤、非硫酸塩系非イオン界面活性剤を投入し,均一に撹拌し,温度を 2℃まで下げて 40-35℃にする;。
校正結果	4)非硫酸塩両性界面活性剤,非硫酸塩非イオン界面活性剤を投入し,均一に撹拌し,40-35℃まで降温する;完成品を取り出す。

[57]中国語の「两」は日本語では個数として「2」と訳すべき場合も多いため、このような置換が行われていると考えられる。実例は後続の事例 3.5.2-3 を参照されたい。

[58] 余った「1」が文末に出力されていない理由については、第 3.5.4 項を参照されたい。

本例の機械和訳は、本来「40-35℃まで降温する」であるべき箇所に「温度を 2℃まで下げて 40-35℃にする」と不要な文言が湧き出している。原文中に算用数字「2」は存在せず、「両性」の「両」から「2」が派生したと考えられる。通常、こうしたケースでは「両性」はそのまま「両性」と訳され、不要な「2」が文末に出力されるが、本例では「両性」が「性」とのみ訳され、「2」は「2℃まで下げて」と不要な文言を伴って文脈に組み込まれるという、複雑かつ重大な不備となっている。サンプル 160 件においても、漢数字から派生した算用数字が同様に文脈に組み込まれ、原文と異なる翻訳文となるケースは 7 件見られた。

以上のように、漢数字から算用数字が派生することに起因する不備の事例は極めて多い。あらゆる文で発生していることから、特定の学習データが原因の現象とは考えにくく、システム上の数値処理が関与していると推察される。そうであれば、この処理を見直すことで根本的に改善される可能性が高い。

3.5.3. 漢数字を含む化合物名称における誤訳

中国語の化合物名称には漢数字が使用されることが非常に多い。前掲の事例 3.5.2-1 や 3.5.2-2 も、化合物名称中の漢数字に起因した不備事例であった。これらの事例では、算用数字の派生による不備は生じたものの、化合物名称自体は正しく訳されていた。だが、化合物名称中の漢数字が算用数字に置換されたことによって誤訳が生じたと見られるケースも多数検出されている。実例を示す。

［事例 3.5.3-1：CN109562045B ｾｸｼｮﾝ A］

原　　文	4.根据权利要求 1 至 3 中任一项所述的组合物，其中所述组合物还包含由选自锌盐、亚锡盐和氯化十六烷基吡啶的口腔可接受的盐提供的口腔可接受的阳离子型活性剂。
機械和訳	4.請求項 1 から 3 のいずれか一項に記載の組成物であって,ここで前記組成物はさらに亜鉛塩、第一スズ塩及び第 16 塩化ジアルキルピリジンから選択される口腔に許容可能な塩により提供される口腔内に許容可能なカチオン性活性剤を含む。
校正結果	4.請求項 1 から 3 のいずれか一項に記載の組成物であって,ここで前記組成物は亜鉛塩,第一スズ塩及び塩化セチルピリジニウムから選択される口腔用として許容可能な塩により提供される口腔用として許容可能なカチオン活性剤をさらに含む。

本例では、化合物名「氯化十六烷基吡啶（塩化セチルピリジニウム）」が機械和訳では「第 16 塩化ジアルキルピリジン」と誤訳されている。語中の漢数字「十六」が「16」に

置換され、名称に組み込まれた形である。

中国対応特許では「氯化十六烷基吡啶」は 214 文の使用例があり、うち 162 文で「塩化セチルピリジニウム」と訳されている。このため、学習データ中でも支配的な訳語であると考えられる。一方、中国対応特許に「第 16 塩化ジアルキルピリジン」という訳例は存在しない。

こうした状況にもかかわらず、本調査対象の請求項で「氯化十六烷基吡啶」が使用されている 13 件（本例を除く）のうち、正しく「塩化セチルピリジニウム」と翻訳されたのは 8 件のみであった。残る 5 件は本例と同様、「16」[59]＋「塩化ジアルキルピリジン」と誤訳されている[60]。

このように、化合物名が漢数字を含む場合、その漢数字が算用数字に置換されたことで学習データの適訳語とマッチせず、誤訳されたと考えられるケースは非常に多い。例えば「三聚氯氰（塩化シアヌル）」→「3 量化クロロシアノ」（CN113789592B）、「聚四氟乙烯（ポリテトラフルオロエチレン）」→「4 フッ化エチレン」（CN112095943B）、「六水合氯化铝（塩化アルミニウム六水和物）」→「6 アルミニウムクロリドロール」（CN113827492B）などがこれに該当する。いずれも、名称中の漢数字が算用数字に置換されて訳語に組み込まれるという共通点がある。

請求項に記載される化合物名称は、発明の根幹に関わる重要な技術用語であることが多い。現状、機械翻訳プラットフォームでは、本項に示したとおり、学習データ中に適訳語が存在するであろう語も、漢数字を算用数字に置換する処理によって誤訳される確率が高くなっている。160 件のサンプルでも、サンプル E2 で「三乙醇胺（トリエタノールアミン）」が「ジエタノールアミン」と誤訳され、さらに漢数字「三」から派生した算用数字「3」が文末に出力されていたり、サンプル D16 では「聚二甲基硅氧烷（ポリジメチルシロキサン）」自体は正しく訳出されたものの、漢数字「二」から派生した「2」が後続の「5~18 モル%」の「5」と置き換わって「2~18 モル%」と誤訳され、あぶれた「5」が文末に出力されていたりと、化合物名称中の漢数字に起因する不備は少なくない（7 件）。この点においても、漢数字の処理方法には改善が求められる。

---

[59] 「16」は文末に別途出力されていたり（CN108951228B、CN108273525B）、「一種又は複数種」の「一」と置き換わって「16 種又は複数種」の形で文脈に組み込まれたり（CN109627848B）している。

[60] なお、正しく「塩化セチルピリジニウム」と訳されているものは、「氯化十六烷基吡啶(CPC)」とカッコ補記を伴うものが 8 件中 6 件と多数を占めた。他方、誤訳されたケースは全てカッコ補記を伴わないものであった。カッコ補記の有無によって翻訳時の扱いが変化している可能性がある。

3.5.4. 漢数字「一」の処理

漢数字「一」は、中国請求項では特に「一种」、「一个」という形で多用される。英語の前置詞（a/an）に近く、日本語では訳出しないことが多い。ここまで本章で示した事例でも、「一种」が事例 3.2.2-2 や事例 3.5.1-2、「一个」が事例 3.3.1-1 で使われていたが、機械和訳、校正結果とも訳出していない[61]。

しかしながら、訳出すべき「一」も当然存在する。そして、調査対象の請求項の中には、訳出すべき「一」が訳出されないケースがいくつか検出された。一例を示す。

［事例 3.5.4-1：CN106962145B ｾｸｼｮﾝ A］

原　　文	所述流量检测单元（7）包括流量传感器一（71）、流量传感器二（72）、流量传感器三（73）和流量传感器四（74），并且流量传感器一（71）、流量传感器二（72）、流量传感器三（73）和流量传感器四（74）的输出端均与微处理器（17）的输入端连接。
	5.根据权利要求 1 所述的一种园林灌溉具有精准控制的水体分流装置，其特征在于：所述指示灯（21）包括指示灯一（211）、指示灯二（212）、指示灯三（213）和指示灯四（214），并且指示灯一（211）、指示灯二（212）、指示灯三（213）和指示灯四（214）的输入端均与中央处理器（19）的输出端连接。
機械和訳	流量検出ユニット(7)は流量センサ(71)、流量センサ 2(72)、流量センサ 3(73)及び流量センサ 4(74)を含み,且つ流量センサ(71)、流量センサ 2(72)、流量センサ 3(73)及び流量センサ 4(74)の出力端はいずれもマイクロプロセッサ(17)の入力端に接続される。
	5.請求項 1 に記載の庭園灌漑に正確な制御を有する水分流装置であって,その特徴は以下のとおりである:前記表示灯(21)は表示灯(211)、表示灯(2)(212)、表示灯(3)(213)及び表示灯(4)(214)を含み,且つ表示灯(211)、表示灯(2)(212)、表示灯(3)(213)及び表示灯(4)の入力端はいずれも中央処理装置(19)の出力端に接続される。214。
校正結果	流量測定ユニット(7)は流量センサ 1(71),流量センサ 2(72),流量センサ 3(73)及び流量センサ 4(74)を含み,且つ流量センサ 1(71),流量センサ 2(72),流量センサ 3(73)及び流量センサ 4(74)の出力端はいずれもマイクロプロセッサ(17)の入力端に接続される。
	5.請求項 1 に記載の庭園の散水のための正確に制御できる水分流装置で

[61] 事例 3.3.3-1 の③行目では「一种」を「一種」と訳出しているが、これは「至少一种（少なくとも一種の）」というフレーズに用いられているためである。

	あって,その特徴は以下のとおりである:前記表示灯(21)は表示灯 1(211),表示灯 2(212),表示灯 3(213)及び表示灯 4(214)を含み,且つ表示灯 1(211),表示灯 2(212),表示灯 3(213)及び表示灯 4(214)の入力端はいずれも中央プロセッサ(19)の出力端に接続される。

上例原文では 1 文目の「流量传感器（流量センサ)」、2 文目の「指示灯（表示灯)」とも漢数字で一～四の番号が付されている。機械和訳は、二～四はそれぞれ「流量センサ 2～4」「表示灯(2)～(4)」と算用数字に置換しているが、「流量传感器一」と「指示灯一」のみは、それぞれ「流量センサ」「表示灯」と訳され、付番「一」が消失している。中国語の「一」は訳出しない場合が多いことから、数値処理において「一」が特殊な扱いをされていることも考えられる[62]。本例はその副作用の可能性がある。事実、160 件のサンプル調査においても、本例と同じく訳出すべき「一」が欠落する不備や、訳出すべきでない「一」が訳出される不備、すなわち漢数字「一」の扱いに関する不備が計 7 例検出されている。

3.5.5. 事例集

数値処理に関しては、ここまでに取り上げたもの以外にも多種の不備が検出されており、現状の機械翻訳プラットフォームにおける最重点課題のひとつといえる。以下、本調査分析で検出した、数値に関連する特徴的な不備の事例を掲載する。

［事例 3.5.5-1：CN107208158B ｾｸｼｮﾝ C］

原　　文	27.如权利要求 1-3、5、8-12、13-14、17-21 和 23-26 中任一项所述的方法，其中所述方法是多重复路的。
機械和訳	27.請求項 1～23-26 のいずれか 1 項に記載のであって,前記方法は多重反復経路であることを特徴とする方法。1－3,5,8－12,13－14,17－21。
校正結果	27.請求項 1-3,5,8-12,13-14,17-21 及び 23-26 のいずれか一項に記載の方法であって,ここで前記方法は多重化して行うことができる。
解　　説	機械和訳は、複雑に列挙された請求項番号を正しく反映できず、大半が文末に羅列されている。同様の不備は、本案件（CN107208158B）の他の請求項をはじめ多数の請求項で見られる。一方で、同様に請求項番号が列挙されている CN105873515B の請求項 30、CN106414323B の請求項 25、CN107614818B の請求項 28 など、正しく訳出されるケースも同様に多数存在する。さらに、列挙された請求項番号が完全に同一であるのに、一方の請求項では正しく訳出され、他方では不備が生じている

[62] 事例集の事例 3.5.5-7 を参照されたい。

	事例もある（CN105324472B の請求項 23～24）。根本的には数値処理ではなく、翻訳文の生成に関わる問題と考えられる。なお、160 件サンプルにおいては同種の不備は検出されなかった。

［事例 3.5.5-2：CN108697595B ｾｸｼｮﾝ A］

原　　文	所述非反应性有机溶剂选自丁烷、戊烷、己烷、环己烷、庚烷、癸烷、十六烷、甲苯、二甲苯、乙酸乙酯、乙酸丁酯、甲基乙基酮、甲基异丁基酮、1,4-二噁烷、甲基氯、二氯甲烷、氯仿、四氯化碳中的一种或多种,
機械和訳	前記非反応性有機溶剤はブタン類,ペンタン類,ヘキサン類,シクロヘキサン類,ヘプタン類,デカン類,ジオキサン類,トルエン類,キシレン類,酢酸エチル類,酢酸ブチル類,メチルエチルケトン類,メチルイソブチルケトン類,ジオキサン類,塩化メチル類,塩化メチレン類,クロロホルム類,4 塩化炭素のうちの一種又は複数種から選択され,16 種又は 1,4-2 種以上である。
校正結果	前記非反応性有機溶媒はブタン,ペンタン,ヘキサン,シクロヘキサン,ヘプタン,デカン,ヘキサデカン,トルエン,キシレン,酢酸エチル,酢酸ブチル,メチルエチルケトン,メチルイソブチルケトン,1,4-ジオキサン,塩化メチル,塩化メチレン,クロロホルム,四塩化炭素のうちの 1 種又は複数種から選択され,
解　　説	機械和訳に、原文には存在しない「16 種又は 1,4-2 種以上である。」という文言が湧き出している。原文の「十六烷」と「1,4-二噁烷」の漢数字から「16」と「2」がそれぞれ派生し、不要な文言として文に組み込まれたと見られる。事例 3.5.2-3 で漢字から派生した「2」から「2℃まで下げて」という文言が湧き出したのと同種のケースである。なお、「十六烷」と「1,4-二噁烷」はそれぞれ「ヘキサデカン」「1,4-ジオキサン」と訳されるべきだが、機械和訳では双方とも「ジオキサン類」と誤訳されている。

［事例 3.5.5-3：CN108193303B ｾｸｼｮﾝ D］

原　　文	9.根据权利要求 8 所述的亲水聚酯纤维的制备方法，其特征在于，所述聚醚二醇选自聚乙二醇、聚丙二醇或聚丁二醇中的至少一种。
機械和訳	9.請求項 8 に記載の親水性ポリエステル繊維の製造方法であって,その特徴は以下のとおりである:前記ポリエーテルグリコールはポリエチレングリコール,ポリプロピレングリコールまたはポリブチレングリコールから選択される少なくとも 2 種である。
校正結果	9.請求項 8 に記載の親水性ポリエステル繊維の製造方法であって,その特徴は以下のとおりである:前記ポリエーテルグリコールはポリエチレング

	リコール,ポリプロピレングリコールまたはポリブチレングリコールから選択される少なくとも 1 種である。
解　　説	原文には漢数字「二」を含む化合物名が四種存在する。その何れかから「2」が派生したと考えられる。さらに、漢数字「一」を訳出しない傾向（⇒第 3.5.4 項）により、「至少一种（少なくとも 1 種である)」の「1」よりも「2」が優先されたと推察される。

［事例 3.5.5-4：CN108265104B ｾｸｼｮﾝ D］

原　　文	2.根据权利要求 1 所述的构建方法，其特征在于，所述步骤 S11 包括：步骤 S111，将所述组织样品剪碎，得到破碎组织；步骤 S112，采用所述红细胞裂解液对所述破碎组织进行裂解，得到所述裂解产物。
機械和訳	2.請求項 1 に記載の構築方法であって,その特徴は以下のとおりである:前記ステップ S11 は以下を含む:ステップ S11 11 であって,前記組織サンプルを切断して,破砕組織を得る;ステップ S11 12,前記赤色細胞可溶化物を用いて前記破砕組織を溶解し,前記溶解産物を得る。
校正結果	2.請求項 1 に記載の構築方法であって,その特徴は,前記ステップ S11 が以下を含むことである:ステップ S111,前記組織サンプルを切断し,破砕組織を得る;ステップ S112,前記赤血球溶解液を用いて前記破砕組織を溶解し,前記溶解産物を得る。
解　　説	機械和訳では、原文中の「S111」「S112」が、それぞれ「S11 11」「S11 12」と 1 桁増加し、かつ中間にスペースが挿入される形で誤訳されている。「S111」「S112」を含む文では必ずこの誤訳が発生している。なお、「S」を伴わず数字（111、112）のみであれば誤訳されない。 「S211」や「S311」も調べたが、それぞれ「S21 1」「S31 1」と、桁数は増加しないが中間スペースは挿入される。発生メカニズムが推測できないが、特定の条件で必ず発生していることから、数値処理に係る不備である可能性が高い。160 件のサンプルでは同種の不備は検出されなかったが、これは本エラーの発生条件が非常に限定的であり、条件に該当する案件自体が含まれていないためである。

［事例 3.5.5-5：CN107795872B ｾｸｼｮﾝ F］

原　　文	1)测量放线：根据图纸中管道走向及标高，用石灰线分别放出管道、管井走向和位置；在管井位置处开挖圆形井坑，在井坑中心处用石灰线撒上“十”形标识，井坑深度应在图纸设计基础上加 100mm 作为雨水井垫层；
機械和訳	1)測定放出:図面中の管路方向及び標高に基づいて,石灰線を用いてそれぞれ管路、管路の井走向及び位置を放出する;管井戸位置で円形井戸ピット

	を掘削し,井戸ピットの中心に石灰線で"10"字形標識を描き,井戸ピット深さは図面設計上に 100mm を雨井戸マット層として加える;
校正結果	1)測定と墨出し:設計図中の管路の方向及び標高に基づいて,石灰ラインを用いて管路,深井戸の方向及び位置を墨出しする;深井戸の位置で円形ピットを掘削し,ピットの中心で石灰ラインで"十"字形標識を描き,ピットの深さは設計図の設計を基礎として雨水枡緩衝層として 100mm 加算する;
解　　説	原文の漢数字「十」は形状を示すために用いられたが、算用数字に変換され、目的を果たせていない。

［事例 3.5.5-6：CN111576341B ｾｸｼｮﾝ E］

原　　文	所述沥青输送管(6)上设置有沥青辅助输送机构，该沥青辅助输送机构包括一产生压缩空气的空气压缩机(7)和一三通阀(702)，该三通阀(702)设置在沥青泵(104)和缺损修补罩(4)之间的沥青输送管(6)上，剩下的一个接口与空气压缩机(7)的压缩空气出口连接，以利用压缩空气对沥青输送管(6)内的沥青起到辅助输送的作用；
機械和訳	前記アスファルト輸送管(6)にアスファルト補助輸送機構が設置され,該アスファルト補助輸送機構は圧縮空気を生成する空気圧縮機(7)及び 1 3 方弁(702)を含み,該 3 方弁(702)はアスファルトポンプ(104)と欠陥補修カバー(4)との間のアスファルト輸送管(6)に設置され,残りの 1 つのインターフェースは空気圧縮機(7)の圧縮空気出口に接続され,圧縮空気でアスファルト輸送管(6)内のアスファルトを補助輸送する役割を果たす;
校正結果	前記アスファルト搬送管(6)にアスファルト補助搬送機構が設置され,該アスファルト補助搬送機構は圧縮空気を生成する空気圧縮機(7)及び三方弁(702)を含み,該三方弁(702)はアスファルトポンプ(104)と欠陥補修カバー(4)との間のアスファルト搬送管(6)に設置され,残りの一つのポートは空気圧縮機(7)の圧縮空気出口に接続され,圧縮空気でアスファルト搬送管(6)内のアスファルトを補助的に搬送する役割を果たす;
解　　説	原文には漢数字「一」が 3 回出現する。このうち最初の 2 回は訳出不要だが、最後の「一个」は訳出したほうが分かりやすい。機械和訳では、その「一个」はうまく訳出している反面、訳出不要な 2 回目の「一」も「1」と訳出している。 だが、この箇所の原文は「一三通阀」と、前置詞的な「一」と技術用語「三通阀」とで漢数字が連続しており、人間でも誤認しかねない。これを適切に切り分ける処理は非常に難易度が高いと想像される。

［事例 3.5.5-7：CN106996547B ｾｸｼｮﾝ F］

原　　文	2.如权利要求 1 所述的易拆装可调角度的灯具支架，其特征在于，每对所述调节盘中，所述调节盘上的相邻两所述轮齿之间设有定位齿，所述轮齿上设有定位槽，所述定位齿与所述定位槽一一对应配合。
機械和訳	2.請求項 1 に記載の着脱可能な角度調整可能な灯具ホルダであって,その特徴は以下のとおりである:各対の前記調整盤において,前記調整盤における隣接する 2 つの前記歯車歯の間に位置決め歯が設けられ,前記歯車座に位置決め溝が設けられ,前記位置決め歯は前記位置決め溝 1 1 に対応して嵌合される。
校正結果	2.請求項 1 に記載の着脱可能で,角度調整可能な照明器具ホルダであって,その特徴は以下のとおりである:各対の前記調整ディスクにおいて,前記調整ディスク上の隣接する二つの前記歯車の間に位置決め歯が設けられ,前記歯車上に位置決め溝が設けられ,前記位置決め歯が前記位置決め溝に一対一に対応して嵌合される。
解　　説	原文の「一一対応」は校正結果にあるように「一対一に対応して」という意味だが、機械和訳では「位置決め溝 1 1 に対応して」と算用数字に置換され、要素付番のようになっている。 「一一対応」は頻出フレーズであり、「一一」を含む全 2,195 項のうち 1,801 項でこのフレーズが使用されていたが、このうち 1,603 項で「1 1」と誤訳されていた（171 項では「1 対 1」と訳された）。サンプル 160 件においても、このフレーズが用いられた 2 件とも同様の誤訳が生じている。なお、「一一連接」、「一一匹配」、「任一一个（いずれか一つ）」なども同様に高率で誤訳されている。 こうした慣用フレーズは学習データ中にも多数の使用例があり、本例のような誤訳が学習データに準拠しているとは考えにくい。エンジン内部で漢数字を数字に置換する処理が、翻訳処理よりも先行して行われた結果の誤訳である可能性が高い。

3.6. 英数字記号の処理

英数字記号[63]も、数字と同様、プレースホルダー方式で翻訳文に反映させることができる。ただし、英数字記号文字列は数値に比べて意味が特定されるため、プレースホルダーに置換して意味情報を失わせるより、英数字記号のまま学習データを参照するほうが、周辺の用語の訳語選定が的確になると考えられている[64]。英数字記号文字列は、原則、そのまま翻訳文に転記すればよく、翻訳難易度は低い。このため、英数字記号が短い場合はプレースホルダーを使用せず、そのまま翻訳処理するのが一般的である。

ただし、英数字記号文字列が長い場合は、中途で分断されるなど機械翻訳処理時に内容が変化するリスクが高くなるため、意味情報の消失より誤訳の防止を優先し、プレースホルダー方式を採用することが多い。機械翻訳プラットフォームの翻訳文では、長大な英数字記号文字列においても反映ミスはほとんど検出されておらず、一定以上の長さの英数字記号文字列にはプレースホルダー方式が使われている可能性が高い。

特許請求項で多用される英数字記号文字列のうち長大・複雑となる傾向が強いものとして、数式と化合物名が挙げられる。第 2 章（第 2.9 項）では、これらについて一定の条件で該当する文を抽出し、重大な不備の発生率を調査した。その結果、少なくとも英数字記号のみで構成される数式、化合物であれば、特に不備を増加させる要因にはなっておらず、むしろ同等の文長の請求項であれば、数式、化合物を含むもののほうが不備発生率が低くなる傾向が示された。これは、英数字記号文字列は原則そのまま翻訳文に反映させればよく、翻訳の必要がないため難易度が下がることによると考えられる。

本項では、この結果を踏まえ、数式や化合物名称を含む請求項の実例を示し、上記調査結果を検証する。さらに、統計的分析では把握できなかった「英数字記号と中国語文字とが混在する数式や化合物名称」の翻訳状況も確認する。

---

[63] 本報告書では、英語・数字・記号が組み合わされた文字列全般を指す。ただし、数字のみの場合は「数値」と見なし、英数字記号には含めない。

[64] プレースホルダー方式では、原文中の文字列を（意味を有さない）所定の目印に置き換える。学習データの参照は、この置換後に行われる。本来、ニューラル機械翻訳方式では、例えば H_2O を含む文であれば、周辺の語句には H_2O と共起する訳語が選定されるが、H_2O がプレースホルダーに置換されていると、こうした手掛かりが失われ、周辺の語句の訳質が低下する懸念がある。

3.6.1. 数式

請求項中の数式は、本文から独立させた形で書かれる場合と、本文に組み込んで書かれる場合とがある。以下、それぞれの事例を示す。

［事例 3.6.1-1：CN107847107B ｾｸｼｮﾝ H］

原　　文	9.根据权利要求 1 所述的医疗用观察装置，其中
	与所述多个光谱分量对应的所述成像信号是 RGB 信号，并且
	当 R、G 和 B 分别是 R 信号、G 信号和 B 信号的辉度分量，并且 r、g 和 b 是基于所述权重并且与所述 R 信号、所述 G 信号和所述 B 信号分别对应的系数时，基于下面列出的公式来计算用于计算所述评估值的评估信号 L：
	[数学公式 1]
	L＝r×R+g×G+b×B。
機械和訳	9.請求項 1 に記載の医療用観察装置であって,その特徴は以下のとおりである:
	前記複数のスペクトル成分に対応する前記撮像信号は RGB 信号であり,且つ
	R、G 及び B がそれぞれ R 信号、G 信号及び B 信号の輝度成分であり,且つ r、g 及び b が前記重みに基づき且つ前記 R 信号、前記 G 信号及び前記 B 信号にそれぞれ対応する係数である場合,以下に列挙される式に基づいて前記評価値を計算するための評価信号 L を計算する:
	[数式 1]
	L=r×R+g×G+b×B である。

［事例 3.6.1-2：CN112267074B ｾｸｼｮﾝ C］

原　　文	1.一种大功率发动机曲轴用高强韧性贝氏体非调质钢，其特征在于，包括如下重量百分比的化学成分：C：0.32-0.38%，Si：0.63-1.25%，Mn：1.60-1.90%，Cr：0.65-0.85%，Mo：0.08-0.12%，V：0.07-0.15%，Ni：0.10-0.18%，Ti：0.022~0.029%，B：0.0020-0.0030%，Al：0.020~0.030%，P：≤0.010%，S：0.027-0.035%，T.O：≤10ppm，[H]：≤1.0ppm，[N]：65-85ppm，205≤X≤245，其余为 Fe 和不可避免的杂质元素；
	其中，X=25.4(0.003+0.53C)×(1+0.7Si)×(-1.12+5.1Mn)×(1+2.16Cr)×(1+1.74V)×(1+0.364Ni)×(1+3Mo)，式中各元素所指数值=该元素在钢中含量×100；
	：

機械和訳	1.大電力エンジンのクランクシャフト用高靭性ベイナイト非調質鋼であって,その特徴は以下のとおりである:以下の重量百分率の化学成分を含む:C：0.32－0.38%，Si：0.63－1.25%，Mn：1.60－1.90%，Cr：0.65－0.85%，Mo：0.08－0.12%，V：0.07－0.15%，Ni：0.10－0.18%，Ti：0.022~0.029%，B：0.0020－0.0030%，Al：0.020~0.030%，P:≦0.010%，S：0.027－0.035%，T.O:≦10wtppm，[H]:≦1.0ppm，[N]：65－85ppm，205≦X≦245,残部は Fe 及び不可避不純物元素;
	ここで,X=25.4(0.003+0.53C)×(1+0.7Si)×(－1.12+5.1Mn)×(1+2.16Cr)×(1+1.74V)×(1+0.364Ni)×(1+3Mo)であり,式中の各元素の指数値=該元素の鋼鉄中含有量×100 である;
	：

前者が本文から独立した形、後者が本文中に組み込まれた形となる。どちらも数式自体の反映は正確に行われている。英数字記号のみで構成される数式は、式全体を一つの文字列として取り扱い、かつ原文の内容をそのまま転記すればよいため、機械和訳に正確に反映することは比較的容易である。文中に組み込まれた式の配置が不正となる可能性はあるが、それは翻訳文生成の精度の問題である。

3.6.2. 化合物名

化合物名も、全体が英数字記号で記載されたものであれば、数式と同様、機械和訳に正確に反映される。ただし、化合物の場合、英数字記号と中国語（漢字）が組み合わされた形で表記されるものも多い。このような表記において、英数字記号文字列を一語と認識するロジックにより英数字部分と中国語部分とが別トークンに分解され、一つの化合物名と認識されなくなる可能性が考えられた。しかしながら、機械和訳を確認したところ、そのような誤訳はほとんど発生していない。以下、一例を示す。

［事例 3.6.2-1：CN106794157B ｾｸｼｮﾝ A］

原　　文	7.如权利要求 1 所述的液体药物组合物，所述组合物包含：(a)9,10-二甲氧基-2-(2,4,6-三甲基苯基亚氨基)-3-(N-氨基甲酰基-2-氨基乙基)-3,4,6,7-四氢-2H-嘧啶并[6,1-a]异喹啉-4-酮或其药学上可接受的盐的颗粒，其浓度为 0.01mg/mL～30mg/mL；(b)聚氧乙烯二醇脱水山梨糖醇烷基酯，其浓度为 0.1mg/mL～2mg/mL；(c)脱水山梨糖醇烷基酯，其浓度为 0.01mg/mL～0.1mg/mL；(d)第一磷酸盐缓冲剂组分，其浓度为 5mg/mL～10mg/mL；(e)第二磷酸盐缓冲剂组分，其浓度为 5mg/mL～10mg/mL；和(f)张力调节剂，其浓度为 2mg/mL～8mg/mL。

機械和訳	7.請求項 1 に記載の液体医薬組成物であって,前記組成物は以下を含む:(a)9,10-ジメトキシ－2－(2,4,6－トリメチルフェニル基イミノ)-3-(N-カルバモイル-2-アミノエチル)-3,4,6,7-テトラヒドロ-2H-ピリミド[6,1-a]イソキノリン-4-オン又はその医薬的に許容される塩の粒子であって,その濃度が 0.01mg/mL~30mg/mL;(b)ポリオキシエチレングリコールソルビタンアルキルエステルであり,その濃度が 0.1mg/mL~2mg/mL;(c)ソルビタンアルキルエステルであり,その濃度が 5mg/mL~10mg/mL である;及び mg/mL~mg/mL である;及び 5mg/mL~mg/mL である;及び 0.01 2 40.1mg/mL;(d)10mg/mL;(e)(f)2mg/mL~8mg/mL の濃度を有する張力調節剤。
校正結果	7.請求項 1 に記載の液体医薬組成物であって,前記組成物は以下を含む:(a)9,10-ジメトキシ-2-(2,4,6-トリメチルフェニルイミノ)-3-(N-カルバモイル-2-アミノエチル)-3,4,6,7-テトラヒドロ-2H-ピリミド[6,1-a]イソキノリン-4-オン又はその医薬的に許容される塩の粒子であって,その濃度が 0.01mg/mL~30mg/mL であるもの;(b)ポリオキシエチレングリコールソルビタンアルキルエステルであって,その濃度が 0.1mg/mL~2mg/mL であるもの;(c)ソルビタンアルキルエステルであって,その濃度が 0.01mg/mL~0.1mg/mL であるもの;(d)第一リン酸塩緩衝剤成分であって,その濃度が 5mg/mL~10mg/mL であるもの;(e)第二リン酸塩緩衝剤成分であって,その濃度が 5mg/mL~10mg/mL であるもの;及び(f)張力調節剤であって,その濃度が 2mg/mL~8mg/mL であるもの。

本例には、「9,10-二甲氧基-2-(2,4,6-三甲基苯基亚氨基)-3-(N-氨基甲酰基-2-氨基乙基)-3,4,6,7-四氢-2H-嘧啶并[6,1-a]异喹啉-4-酮」という、長大かつ中国語と英数字記号が複雑に組み合わされた化合物名が記載されている。こうした語に対して、英数字記号部分と中国語部分とを分離して個別に扱った場合、化合物名称中の語順や文全体のトークン分割などに支障が出る可能性が高いと考えた。しかしながら、本例の機械和訳を見ると、途中に不要な「基」はあるものの、化合物の語順は原文どおり正確に再現されており、翻訳文全体の翻訳精度にも悪影響は見られない。

もう一例示す。

［事例 3.6.2-2：CN113622052B ｾｸｼｮﾝ D］

原　　文	7.根据权利要求 6 所述的阻燃单丝的制备方法，其特征在于，所述第二阻燃剂为乙基磷酸二乙酯；所述柔软剂选自间苯二酚双(磷酸二苯酯)和双酚 A-双(磷酸二苯酯)中的至少一种；所述第二交联剂位含环氧基团的多功能聚合物；所述抗氧剂为质量比为 2：3 的 β-(3，5-二叔丁基-4-羟基苯基)丙酸正十八碳醇酯和双(2，4-(二苯代)叔丁基苯基)季戊四醇撑双亚磷酸酯的混合物。
機械和訳	7.請求項 6 に記載の難燃性モノフィラメントの製造方法であって,その特徴は以下のとおりである:前記第二難燃剤はエチルリン酸ジエチルである;前記柔軟剤はレゾルシノールビス(ジフェニルホスフェート)及びビスフェノール A-ビス(ジフェニルホスフェート)からなる群から選択される少なくとも 1 種である;前記第二架橋剤ビットはエポキシ基含有多機能ポリマーを含む;前記酸化防止剤は質量比が 2:1 である:2 23 の β-(3,5-2-tert-ブチル-4-ヒドロキシフェニル)プロピオン酸 n-ヘキサノールとビス(2,4-(ジフェニル)tert-ブチルフェニル)ペンタエリスリトールビスリン酸エステルの混合物。18 2。
校正結果	7.請求項 6 に記載の難燃性モノフィラメントの製造方法であって,その特徴は以下のとおりである:前記第二難燃剤はエチルホスホン酸ジエチルである;前記柔軟剤はレゾルシノールビス(ジフェニルホスフェート)及びビスフェノール A-ビス(ジフェニルホスフェート)から選択される少なくとも 1 種である;前記第二架橋剤はエポキシ基含有多機能ポリマーを含む;前記酸化防止剤は質量比が 2:3 の β-(3,5-ジ-tert-ブチル-4-ヒドロキシフェニル)プロピオン酸オクタデシルとビス(2,4-(スチルベン)tert-ブチルフェニル)ペンタエリスリトールビスホスフィットの混合物である。

本例にも中国語と英数字記号が組み合わされた化合物名が 3 種記載されている。それぞれの機械和訳と校正結果を一覧表にした。

①	双酚 A-双(磷酸二苯酯)	
	機械和訳	ビスフェノール A-ビス(ジフェニルホスフェート)
	校正結果	ビスフェノール A-ビス(ジフェニルホスフェート)
②	β-(3，5-二叔丁基-4-羟基苯基)丙酸正十八碳醇酯	
	機械和訳	β-(3,5-2-tert-ブチル-4-ヒドロキシフェニル)プロピオン酸 n-ヘキサノール

	校正結果	β-(3,5-ジ-tert-ブチル-4-ヒドロキシフェニル)プロピオン酸オクタデシル
③	(2，4-(二苯代)叔丁基苯基)季戊四醇撑双亚磷酸酯	
	機械和訳	ビス(2,4-(ジフェニル)tert-ブチルフェニル)ペンタエリスリトールビスリン酸エステル
	校正結果	ビス(2,4-(スチルベン)tert-ブチルフェニル)ペンタエリスリトールビスホスフィット

これら 3 つの化合物名の機械和訳を見ると、②と③で化合物名を構成している化学物質名がいくつか誤訳されており、中国語部分の翻訳精度には課題もある。ただし、各化合物名は正しい単位で扱われており、英数字部分と中国語部分が分離して語順が乱れていたり、化合物間で語が入り混じったりはしていない。

これらの事例に代表されるように、機械翻訳プラットフォームは、中国語と英数字記号とが組み合わされた化合物名の取り扱いに関しても、妥当に対処できている。160 件のサンプルにおいても、数式、化合物名称など英数字記号を含む文字列に起因する不備は検出されなかった。これらの調査結果から、機械翻訳プラットフォームは英数字記号関連の処理に関しては十分な精度を有していると結論できる。

3.6.3. 事例集

本調査分析で検出した、英字記号文字列の処理に関連する事例をいくつか掲載する。

［事例 3.6.3-1：CN109667023B ｾｸｼｮﾝ D］

原　　文	（5）接结组织在面料的表面形成方格形图案，方格的边长由一花的经、纬向循环数决定，经纬向循环数越大，方格也就越大；
	本组织设计时，方格和边长设定为 2.5cm，2 个方框为一个完整的经向循环；
	表层经纱的密度为 256 根/10cm，表层经纱的循环数=256 根/10cm*5cm/10=128 根，表层组织为平纹，因此一花循环为 64 个循环；
	表、里层一花总的循环数为 256 根；
	纬纱的排列比为 1:1:1，每层的纬纱循环数为 6 纬，表层纬密为 228 根/10cm，为使表层形成方格图案，表层组织的纬纱的循环数=表层纬密/10*纬向一花宽度=22.8 根/cm*2.5cm=57 根纱；
	表层组织的纬纱循环数为 2，因此表层组织的循环个数为 28 或 29 个循环，然后加一根接结纬；
	本面料选用 28 个循环+一根接结纬，纬向基础组织循环数为 6 纬，纬向总

	循环数为 28*6+一根接结纬，共计 169 纬；
機械和訳	(5)接合組織は生地の表面に格子状パターンを形成し,碁盤目の辺長は花の経、緯周期の循環数によって決定され,緯周期循環数が大きいほど,格子も大きくなる;
	本組織の設計時に,格子と辺の長さは 2.5cm に設定され,2 つのブロックは一つの完全な経方向循環である;
	表層経糸の密度は 256 本/10cm であり,表層経糸の循環数=256 本/10cm*5cm/10=128 根であり,表層組織は平織であるため,一花サイクルは 64 サイクルである;
	表、裏層の総サイクル数は 256 本である;
	緯糸の配列比が 1:1:1 であり,各層の緯糸サイクル数が 6 ピックであり,表層の横糸が 228 本/10cm であり,表層に格子模様を形成させ,表層組織の緯糸の循環数=表層の横糸/10*横糸の幅=22.8 本/本の継ぎ目;
	表層組織の横糸循環数は 2 であり,従って表層組織の循環数は 28 又は 29 サイクルであり,その後に一つの接合緯入れを行う;cm*2.5cm=57
	本生地は 28 つの循環+1 本の結合緯入れを選択し,緯側基礎組織循環数は 6 本であり,緯側総循環数は 28*6+1 本の接合緯入れであり,合計 169 本である;
校正結果	(5)接合組織は生地の表面に格子状パターンを形成し,碁盤目の辺長は花の経,緯方向のループ数によって決定され,経,緯方向のループ数が大きいほど,格子も大きくなる;
	本組織の設計時に,格子と辺の長さは 2.5cm に設定され,2 つのブロックは一つの完全な経方向ループである;
	表層経糸の密度は 256 本/10cm であり,表層経糸のループ数=256 本/10cm*5cm/10=128 本であり,表層組織は平織であるため,一花サイクルは 64 サイクルである;
	表,裏層の一花総サイクル数は 256 本である;
	緯糸の配列比が 1:1:1 であり,各層の緯糸ループ数が 6 ピックであり,表層の緯密が 228 本/10cm であり,表層に格子状パターンを形成させ,表層組織の緯糸のループ数=表層の緯密/10*緯方向一花の幅=22.8 本/cm*2.5cm=57 本糸である;
	表層組織の緯糸ループ数は 2 であり,従って表層組織のループ個数は 28 又は 29 個ループであり,その後に一本の接合緯入れを行う;
	本生地は 28 個のループ+1 本の結合緯入れを選択し,緯方向基礎組織ループ数は 6 緯であり,緯方向総ループ数は 28*6+1 本の接合緯入れであり,合計 169 緯である;

解　　説	本例には中国語が組み込まれた数式が三つ含まれる（3,5,7 行目）。機械翻訳処理では、中国語部分と英数字記号部分とを切り離し、中国語部分のみを翻訳し、英数字記号部分はそのまま転記していると考えられる。3 行目と 7 行目の式は正しい構造で反映されている。 一方、5 行目では、式の末尾の英数字記号部分「cm*2.5cm=57」を式中に組み込めず、6 行目の行末に出力されている[65]。これは、中国語が含まれることで数式全体を一つの文字列として取り扱えなかったことによるトークン分割の失敗によるものと考えられる。ただし、式中の英数字記号部分の転記自体は正確に行われている。

［事例 3.6.3-2：CN105441384B ｾｸｼｮﾝ C］

原　　文	2.权利要求 1 所述的方法，其中所述 MAPK 通路抑制因子包括 AEG 3482,BI 78D3,CEP1347,c-JUN peptide,IQ 3,JIP-1(153-163),SP 600125,SU 3327,TCS JNK 5a,TCS JNK6o,AMG 548,CMPD-1,EO 1428,JX 401,ML 3403,RWJ 67657,SB 202190,SB 203580,SB203580hydrochloride,SB 239063,SCIO 469hydrochloride,SKF 86002dihydrochloride,SX 011,TAK 715,VX 702,VX 745,Arctigenin,10Z-Hymenialdisine,PD 0325901,PD184352,PD 198306,PD 334581,PD 98059,SL 327,U0124 和 U0126 中的任意一种或多种。
機械和訳	2.請求項 1 に記載の方法であって,ここで前記 MAPK 経路抑制因子は AEG 3482,BI 78D3,CEP1347,c－JUN peptide,IQ 3,JIP－1(153－163),SP 600125,SU 3327,TCS JNK 5a,TCS JNK6o,AMG 548,CMPD－1,EO 1428,JX 401,ML 3403,RWJ 67657,SB 202190,SB 203580,SB203580hydrochloride,SB 239063,SCIO 469hydrochloride,SKF 86002dihydrochloride,SX 011,TAK 715,VX 702,VX 745,Arctigenin,10Z－Hymenialdisine,PD 0325901,PD184352,PD 198306,PD 334581,PD 98059,SL 327,U0124 及び U01 26 のいずれか 1 つ又は複数を含む。
解　　説	英数字記号文字列がカンマ区切りで羅列され、機械翻訳が不得手な長文（全 410 文字）となっているが、英数字記号は機械和訳に正確に転記されている。英数字記号文字列の反映の安定性が示された事例である。 ニューラル機械翻訳では、極端に長い英数字文字列はプレースホルダー方式で取り扱う場合があり、機械翻訳プラットフォームのエンジンもそうである可能性がある。

[65] この結果から、6 行目と 7 行目はセミコロンで区切られず、同じ翻訳処理単位で処理されたと考えられる。

3.7. 請求項の記載様式に係る不備

本項では、中国請求項の記載様式に係る特徴的な不備を分析する。

3.7.1. 請求項番号が本文の一部として扱われる不備

中国特許公報の請求項は、冒頭に「半角数字＋半角ピリオド」（例：「1.」「2.」）のフォーマットで請求項番号が記載される。

通常、請求項番号は機械和訳にそのまま反映される。だが、一部の請求項では、請求項番号が本文の一部と誤認され、誤った形で本文に組み込まれていた。以下、典型的な事例を二種示す。

［事例 3.7.1-1：CN108180408B ｾｸｼｮﾝ F］

原　　文	1.一种中继镜头，其特征在于，所述中继镜头包括第一部分和第二部分；所述第一部分的第一入射面为与光路垂直的平面；所述第一部分的第一出射面为凸球面；所述第二部分位于所述第一部分的中心位置；所述第二部分的第二入射面为与所述光路垂直的平面；所述第二部分的第二出射面的曲率小于所述第一部分的第一出射面的曲率；所述第二出射面的最大截面尺寸，通过如下方式得出：计算不同的所述第二出射面的截面尺寸与照明效率及照明均匀性的对应关系；选择所述照明效率和照明均匀性中的任一项达到设定最低值时对应的尺寸，作为所述第二出射面的最大截面尺寸。
機械和訳	請求項 1 に記載の方法において,前記リレーレンズを第一部分と第二部分とを具備し,前記第一の部分の第一の入射面は,光路と垂直な平面であり,前記第一の部分の第一の出射面が凸球面である前記第二の部分は前記第一部分の中心位置に位置しており,前記第二の部分の第二の入射面が,前記光路と垂直な平面であり,前記第二の部分の第二の出射面の曲率が前記第一の部分の第一の出射面の曲率よりも小さく,前記第二の出射面の最大断面寸法は,前記第二出射面の異なる断面サイズと光効率と照明均一度との対応関係を算出し,前記照明効率および照度の均一性のいずれかに設定され,最も低い値に対応する大きさに到達するように選択し,前記第二の出射面の最大断面寸法であることを特徴とするリレーレンズである。
校正結果	1.リレーレンズであって,その特徴は以下のとおりである:前記リレーレンズは第一部分と第二部分を含み,前記第一部分の第一入射面は光路と垂直な平面であり,前記第一部分の第一出射面は凸球面である;前記第二部分は前記第一部分の中心位置に位置し,前記第二部分の第二入射面は前記光路と垂直な平面であり,前記第二部分の第二出射面の曲率は前記第一部分の第一出射面の曲率よりも小さい;前記第二出射面の最大断面寸法は以下の

	方式で取得される:前記第二出射面の断面サイズごとの照明効率と照明均一性との対応関係を算出する;前記照明効率および照明均一性のいずれか一つが設定された最低値に達した時の対応する寸法を選択し,前記第二出射面の最大断面寸法とする。

［事例 3.7.1-2：CN107665903B ｾｸｼｮﾝ H］

原　　文	1.一种有机发光显示装置，包括：具有多个子像素的基板；分别设置在所述多个子像素中的多个第一电极；设置在每个所述第一电极的边缘上并与所述第一电极接触的第一电极保护图案；配置成与所述第一电极保护图案的一部分交叠以限定发光区域的堤部；位于所述第一电极上以对应于所述发光区域的有机发光层；以及位于所述有机发光层上的第二电极，其中，所述堤部包括：第一层，所述第一层配置成与每个所述第一电极的一部分交叠，使得相邻子像素的每个第一电极之间的区域被所述第一层填充；和第二层，所述第二层位于所述第一层上以对应于相邻子像素的每个第一电极之间的区域，以及其中，所述第一层是亲水绝缘材料并且所述第二层是疏水绝缘材料。
機械和訳	図 1 は,複数の副画素を有する基板と,前記複数のサブ画素のそれぞれに設けられた複数の第一の電極と,第一の電極の縁部上に配置され,前記第一の電極に接触する第一の電極パターンと,を含み,前記第一電極パターンの一部と重なることを特徴とする発光領域を区画するバンクに配置され;前記第一電極上には,前記発光領域の有機発光層であることを特徴とする有機発光表示装置。前記有機発光層の上に位置する第二電極ここで,前記堤部は以下を含む:第一層であって,前記第一層は各前記第一電極の一部と重なり合うように配置され,隣接するサブ画素の各第一電極間の領域が前記第一層によって充填されるもの;そして第二層であって,前記第二層は前記第一層の上に位置し隣接するサブ画素の各第一電極の間の領域に対応し,及びここで,前記第一層は親水性絶縁材料であり且つ前記第二層は疎水性絶縁材料である。
校正結果	1.有機発光ディスプレイであって,以下を含む:複数のサブ画素を有する基板;それぞれ前記複数のサブ画素に設けられた複数の第一電極;前記第一電極の各々の縁部上に設置され,前記第一電極に接触する第一の電極パターン;前記第一電極保護パターンの一部と重なり発光領域を区画するように構成されるバンク;前記第一電極上に位置し前記発光領域に対応する有機発光層;および前記有機発光層上に位置する第二電極であって,ここで,前記バンクは以下を含む:第一層であって,前記第一層は各前記第一電極の一部と重なり合い,隣接するサブ画素の各第一電極間の領域が前記第一層に

	よって充填されるように構成される;および第二層であって,前記第二層は前記第一層上に位置し隣接するサブ画素の各第一電極の間の領域に対応し,及びここで,前記第一層は親水性絶縁材料であり且つ前記第二層は疎水性絶縁材料である。

これら二例では、原文冒頭の請求項番号「1.」が、それぞれ「請求項 1 に記載の方法であって」、「図 1 は,」と誤った解釈に基づいて機械和訳の本文に組み込まれている。こうした誤解釈には多数のバリエーションが見られるが、これら二種は特に頻出であった。

請求項番号が本文中に組み込まれるこうした不備は、請求項 2 以降でも同様に発生している。下表に請求項 1〜5 における発生頻度を示す[66]。

	全請求項		不備発生請求項		
	請求項数	平均文長	請求項数	発生率	平均文長
請求項 1	100,000	460.2 文字	4,455	4.5%	519.1 文字
請求項 2	95,777	107.0 文字	274	0.3%	320.1 文字
請求項 3	91,953	102.6 文字	247	0.3%	302.9 文字
請求項 4	86,914	104.7 文字	274	0.3%	318.5 文字
請求項 5	80,575	106.0 文字	278	0.3%	316.1 文字

不備は各請求項で発生しているが、請求項 1 の発生率が特に高く、請求項 2 以降では長大な請求項に偏って発生している。このことから、この不備は長大な請求項で発生する傾向が強いことがわかる。請求項 1 での発生率が高いのは、請求項 1 が常に独立請求項であり、概して長大なためと考えられる。160 件サンプルでは 13 件で同種の不備が検出されたが、うち 9 件が請求項 1 であり、残る 4 件（請求項 2 と請求項 5 で各 2 件）もそれぞれ 341 文字、621 文字、401 文字、181 文字と長大な請求項での発生が多かった。

請求項番号が原文と無関係な文言とともに翻訳文に組み込まれるというこの不備の状況は、第 3.5.2 項の事例 3.5.2-3 で見られた数値処理の不備、すなわち漢字「両」から派生した「2」が「2℃まで下げて」という無関係な文言を伴って翻訳文に組み込まれた状況と似ている。このことから、機械翻訳プラットフォームのエンジンは、請求項番号も本文中の数値と同様に扱っていることが窺える。

[66] 請求項 1 であれば『原文の冒頭が「1.」であり、かつ機械和訳の冒頭が「1.」でない』請求項をカウントする形で計測した。

こうした状況から、この不備の発生原因は、入力文が長大なことによる翻訳処理精度の低下と、数値処理上の問題であると考えられる。その場合、ユーザが学習データを通じて解消することは困難である。ただし、発生条件が特定されるため、開発者にとっては比較的対処しやすい不備である可能性がある。不要な情報が文脈に組み込まれるため、正確な内容理解への悪影響も大きく、優先的に解決すべき不備の一つに挙げられる。

具体的な対処策としては、文頭の「数字＋ピリオド」は前処理で本文と切り離し、プレースホルダー化や機械翻訳処理を迂回させて、後処理で翻訳文の文頭に戻すという方法が考えられる。ただし、その際は「1.1」などの箇条番号が付された文や、数値から始まる自然文などを正しく見分け、適切に対処する工夫が必要となる。

3.7.2. 発明の主題を文末に出力する傾向

中国と日本の請求項は記載様式に違いがある。中国の請求項は発明の主題を冒頭で述べるのに対し、日本の請求項では文末に配置する傾向がある。以下、中日対応特許の請求項で例示する。

［参考］中国公開公報 CN101126033A「特許請求の範囲」対訳文

中国請求項	1.一种乳剂燃料，其特征在于，含有燃料油和功能水。
	2.如权利要求 1 所述的乳剂燃料，其特征在于，功能水为电子水和高谐波还原水。
	3.如权利要求 1 或 2 所述的乳剂燃料，其特征在于，还含有氯化钠。
日本請求項	燃料油と機能水とからなることを特徴とする、エマルジョン燃料。
	請求項 1 記載のエマルジョン燃料において、機能水は電子水及び高調波還元水であることを特徴とする、エマルジョン燃料。
	請求項 1 または 2 記載のエマルジョン燃料において、更に塩化ナトリウムを含むことを特徴とする、エマルジョン燃料。

中国請求項では、発明の主題「乳剂燃料」は各項とも文の冒頭に記載されている。これに対し、日本請求項は、請求項 1 では「エマルジョン燃料」は文末に配置され、請求項 2 と 3 では冒頭と文末に 2 回繰り返されている。

中日対応特許由来の対訳コーパスには、このように様式の異なる中日の請求項を対訳文としたデータが大量に含まれる。こうしたデータを学習したことで、機械翻訳プラットフォームの和訳文には、発明の主題を文末に出力しようとする傾向が生じている。サンプル

160 案件においても 29 件にこの傾向が見られる請求項が存在した。

この傾向自体は必ずしも不備ではなく、請求項が短文で、冒頭から文末まで一括で翻訳処理される場合には問題とならない。だが、長大な請求項を複数に分割して翻訳した場合は、原文で発明の主題が書かれた箇所（つまり冒頭部）と、翻訳文で発明の主題を出力しようとする箇所（つまり末尾部）とが、別々の翻訳処理単位となることで問題が生じる。現状の機械翻訳では、翻訳処理単位間で情報を受け渡すことができないためである。一例を示す。

［事例 3.7.2-1：CN107920803B ｾｸｼｮﾝ A］

原　　文	1.一种超声波阵列振荡器，包括：基板；
	彼此独立的多个超声波振荡器模块，所述多个超声波振荡器模块构成阵列；
	其中所述多个超声波振荡器模块的每一个包含：超声波振荡器；
	和半导体芯片，所述半导体芯片形成用于所述超声波振荡器的阻抗匹配电路，所述半导体芯片接合到所述超声波振荡器并设置在所述基板和所述超声波振荡器之间，并且所述半导体芯片将设置在所述基板上的布线电连接至所述超声波振荡器。
機械和訳	1.超音波アレイ発振器であって,以下を含む:基板;
	互いに独立した複数の超音波振動子モジュールであって,前記複数の超音波振動子モジュールがアレイを構成するもの;
	ここで前記複数の超音波振動子モジュールの各々は以下を含む:超音波振動子;
	そして超音波振動子用インピーダンス整合回路を形成する半導体チップであって,前記半導体チップは前記超音波振動子に接合されて前記基板と前記超音波振動子との間に配置され,前記半導体チップは,前記基板上に配置された配線を前記超音波振動子に電気的に接続することを特徴とする半導体チップ。
校正結果	1.超音波アレイ振動子であって,以下を含む:基板;
	互いに独立した複数の超音波振動子モジュールであって,前記複数の超音波振動子モジュールがアレイを構成するもの;
	ここで前記複数の超音波振動子モジュールの各々は以下を含む:超音波振動子;
	および超音波振動子用インピーダンス整合回路を形成する半導体チップであって,前記半導体チップは前記超音波振動子に接合され,且つ前記基板と前記超音波振動子との間に配置され,且つ前記半導体チップは,前記基板

	上に配置された配線を前記超音波振動子に電気的に接続するもの。

本例の発明の主題は、1 文目の「超声波阵列振荡器（超音波アレイ振動子)」である。したがって、日本様式で訳すならば、1 文目冒頭が「1.超音波アレイ振動子であって,」であれば、4 文目末尾も「…ことを特徴とする超音波アレイ振動子。」と同じ主題を繰り返す必要がある。だが、本請求項は意図的改行で 4 文に区切られており、1 文目と 4 文目とは別の翻訳単位として処理される。このため、1 文目に存在する発明の主題を 4 文目の翻訳処理に引き継ぐことはできない。その一方で、機械翻訳エンジンは、請求項の文末に発明の主題を配置するよう学習されている。その結果、本例では、4 文目の主語である「半導体チップ」という一構成要素があたかも発明の主題であるかのように文末にも出力され、読み手に違和感を抱かせる。

ただし、本例に関しては、このことで発明の理解が大きく妨げられるとまではいえない。発明の一構成要素がこの行のみで強調される点に不自然さはあるが、4 文目を「半導体チップ」の説明文と捉えれば、内容的には間違いはない。

長大な請求項は、発明の構成要素の単位で文が区切られているものが多い。このため、一構成要素が文末で繰り返されても、本例と同様、内容理解に致命的な問題は生じないケースが大半と考えられる。ただし、請求項の構成によっては、この傾向が内容を不正確にする場合もある。一例を示す。

［事例 3.7.2-2：CN106394270B ｾｸｼｮﾝ B］

原　　文	1.一种车辆无线充电系统，包括：
	逆变器；
	至少一个控制器，被配置为：
	根据到车辆电力转换器的电压输入与电压基准之间的误差改变所述逆变器的输出电压的频率，以控制所述电压输入，从而使得所述逆变器的输出处的阻抗相位角趋向于预定角度并达到车辆电力转换器的输出处的功率需求，并且基于加权的目标函数的变化率改变所述逆变器的输出电压的频率，其中，所述电压基准是基于所述功率需求、车辆电力转换器的功率输出、所述阻抗相位角以及所述预定角度的，所述加权的目标函数被配置为减小车辆电力转换器的输出功率误差和所述逆变器的输出处的阻抗相位角误差。
機械和訳	1.車両無線充電システムであって,以下を含む:
	インバータ;
	少なくとも一つのコントローラであって,以下のように構成されるもの:

	前記インバータの出力電圧の周波数は,前記インバータの出力におけるインピーダンス位相角が所定角度になるように前記インバータの出力電圧の周波数を変化させて車両電力コンバータ出力の電力需要に到達するように車両電力変換装置の電圧入力と電圧基準との誤差に応じて変更され,前記電力要求,自動車電力変換装置の電力出力,前記インピーダンス位相角及び前記所定角度に基づいて,車両電力コンバータ出力電力誤差及び前記インバータ出力のインピーダンス位相角誤差を低減するように構成されることを特徴とする車両電力変換装置。
校正結果	1.車両無線充電システムであって,以下を含む:
	インバータ;
	少なくとも一つのコントローラであって,以下のように構成される:
	車両電力変換器への電圧入力と電圧基準との誤差に基づき前記インバータの出力電圧の周波数を変えて,前記電圧の入力を制御し,それにより前記インバータの出力におけるインピーダンスの位相角が所定の角度に近づき,かつ車両電力変換器の出力における電力需要に達し,かつ加重目的関数の変化率に基づき前記インバータの出力電圧の周波数を変更し,ここで前記電圧基準は前記電力要件,車両の電力変換器の電力出力,前記インピーダンス位相角および所定の角度に基づくもので,前記加重目的関数は車両の電力変更器の出力電力の誤差および前記インバータの出力におけるインピーダンス位相角の誤差を低減するように構成される。

本請求項も意図的改行で 4 文に分割されており、1 文目と 4 文目は別々に翻訳処理されている。発明の主題は 1 文目の「车辆无线充电系统（車両無線充電システム）」だが、4 文目にはこの情報を引き継げず、機械翻訳エンジンは、最終文の範囲から「車両電力変換装置」を主題と判断し、文末に重複出力した。

しかしながら、本例の 4 文目は、3 文目に挙げられた発明構成要素「コントローラ」の説明文である。このため、4 文目の主題は「コントローラ」としなければならない。だが、4 文目には「コントローラ（控制器）」という語自体が出現せず、文中のどの語句を主題としても誤りとなる。本例のような構成の請求項では、主題を文末で繰り返すこと自体が誤訳の直接の原因となる。サンプル 160 件中、主題の文末配置が見られた 29 件においても、本例と同様の形で重大な不備が発生したケースが 7 件存在していた。

主題を文末に重複出力する傾向は、中日対応特許由来のコーパスを学習データに用いる場合、必然的に生じると考えられる。ただし、その場合も、多くは事例 3.7.2-1 のように違和感を生む程度で、発明の内容理解への致命的な悪影響はない。このため、最優先に解

決すべき課題とはいえない。

仮にこの課題に対処するとすれば、学習データから、請求項部分に由来する文例を一括除去することが有効であろう。その結果、学習データの一部が失われることになるが、請求項に記載された文言や内容の大半は明細書本文でも述べられており、実質的に失われる情報は限定的と考えられる。明細書本文の文体は中国公報、日本公報とも基本的に自然文であり、請求項のような記載様式の違いがない。このため、対訳データとしての精度が高く、こうしたデータのみに絞ることで、全体の翻訳品質にむしろ好影響を与える可能性もある。

3.7.3. 発明構成要素の訳ゆれ

訳ゆれとは、同一の用語が、一請求項内や一案件内において複数の訳語に訳されている状態を指す。一例を示す。

［事例 3.7.3-1：CN107074359B ｾｸｼｮﾝ B］

原　　文	2.根据权利要求 1 所述的航空器，其中所述止挡构件被配置成使所述推进器通过以下动作中的至少一者停止：从所述马达移除电流；使至所述马达的所述电流的极性逆转；将止挡杆部署至所述马达的转子中；或使所述推进器从所述马达脱离。
機械和訳	2.請求項 1 に記載の航空機であって,ここで前記ストッパ部材は前記スラスタが以下の動作のうちの少なくとも 1 つによって停止するように配置される:前記モータから電流を除去する;前記モータへの前記電流の極性を反転させる;ストッパロッドを前記モータの回転子に配置する;又は前記推進器を前記モータから離脱させる。
校正結果	2.請求項 1 に記載の航空機であって,ここで前記ストッパ部材は前記プロペラが以下の動作のうちの少なくとも一つによって停止するように配置される:前記モータから電流を除去する;前記モータへの前記電流の極性を反転させる;ストッパロッドを前記モータの回転子に配置する;又は前記プロペラを前記モータから離脱させる。

本件は、原文中に二度出現する「推进器」が、機械翻訳では「スラスタ」と「推進器」に訳ゆれしている。このため、一読して両者が同じ構成要素を指すとは認識し難い。

本件は自動航空機（ドローン）の発明であり、「推进器」は「プロペラ」を訳すのが適切である。だが、本件の全請求項で「推进器」は 37 回出現するが、機械和訳で「プロペラ」と訳されたのは 4 回のみであり、あとは「スラスタ」が 13 回、「推進器」が 10 回、

「プッシャ」が 6 回、「推進装置」が 2 回、「インペラ」が 1 回と、一案件中で 6 種の訳語に訳ゆれしていた[67]。

ニューラル機械翻訳で訳ゆれが発生する原因は、学習データ中に一つの用語に対して複数の訳語候補が存在することにある。対応特許由来のコーパスは、不特定多数の出願人の手による特許公報をソースとする。技術用語の多くは、個々の公報内では用語統一されていても公報間では統一されていない。このため、学習データ中に複数の訳語候補が併存することとなる。

そして、ニューラル機械翻訳は、同一の用語であっても、周辺語句の状況等に応じ、学習データ中から最も適切な（≒ベクトル値が近似する）訳語をその都度選択する。これにより技術分野や文脈に応じた適切な訳し分けが可能となるが、その反面、同一の用語に対して異なる訳語が採用される可能性も常にはらむ。

用語の訳ゆれは、内容理解を全く阻害しない場合もある。例えば動詞や一般用語は、訳ゆれしていても読み手は混乱しない。例えば「设置」が「設ける」と「設置する」に訳ゆれしたとしても、発明内容を理解する上で問題は生じない。

だが、技術用語、とりわけ特許請求の範囲における発明の構成要素は、全請求項で統一名称で記載されるのが共通認識となっており、訳ゆれが発生すると、読み手の読解を妨げ、場合によってはそれぞれが別個の要素と誤認されて、発明の内容理解に支障をきたす。本事例はその典型例である。

訳ゆれは文中や案件中に同一の用語が複数回出現することが発生条件であり、用語の繰り返しが多いほど発生確率が上がる。特許請求の範囲は、請求項内や請求項間で各構成要素が繰り返し言及される傾向が強く、訳ゆれが発生しやすい文書といえる。160 件のサンプリングでも、発明の構成要素など訳語統一されるべき語句が訳ゆれしていた案件は 53 件と多発した[68]。また、サンプル A3 では「枢转销」→「ピボットピン」「枢動ピン」「回動ピン」「旋回ピン」、A4 では「打浆容器」→「叩解容器」「パルプ容器」「スラリー容器」など、案件中で繰り返し使われる語句が 3 種以上の訳語に訳ゆれするケースも目立った。

---

[67] 残る 1 回は訳抜け。

[68] 訳ゆれにより、第 3.2.2 項で述べた［訳語候補が複数存在することによる誤訳］となったものはそちらにカウントした。ここでカウントされた 53 件は、それらを除く「訳ゆれしているが、いずれも同義語の範疇にある」ものとなる。

用語の訳ゆれは、学習データからその都度最適の訳語を採用するニューラル機械翻訳方式においては宿命的な現象といえる。現状のニューラル機械翻訳で特定の語の訳語を統一するには、学習データから他の訳語を全て排除する必要があり、現実的ではない。

一部のニューラル機械翻訳エンジンでは、学習データより優先的に参照する辞書を使用することができる。そこに訳語を登録することで、訳語を固定させることが可能である。だが、そもそも訳ゆれが発生する技術用語とは、学習データ中に複数の有力な訳語が存在する用語であり、例えば事例 3.2.2-2 の「上网（網／インターネット）」のように、分野や文脈によって適訳語が異なるものも多い。こうした語は、辞書登録で訳語を一意に固定すると、訳ゆれは解消される反面、分野・文脈に応じた適切な訳し分けというメリットが失われ、相当量の不適訳が確実に発生する結果となる。

このため、現状では、請求項内や請求項間での用語の訳ゆれは解決困難である。ただし、ニューラル機械翻訳分野では、この課題に対処すべく、前後の文の文脈を考慮して訳語の統一を図る手法の研究が進められている。将来的には、請求項単位や案件単位で訳語情報を共有し、訳語の統一が達成されることが期待される。

3.7.4. 長大な請求項での翻訳精度の低下

請求項は原則一文で記述される。このため、発明の全体像が詳細に記述される独立請求項は、一般的な文に比べて長大となる傾向が強い。本調査対象の独立請求項全件の平均文長は 420.4 文字であった。

文は、長文となるほど構造が複雑になり、語彙も増加していく。このため機械翻訳の難易度は高くなる。第 2 章の統計的調査分析でも、文長と校正率や重大不備の発生率との間には明らかな相関が見られた。

一文が長大となることが特許請求項の本質的な特徴である以上、長文での翻訳精度の改善は、機械翻訳プラットフォームが果たすべき重要な課題といえる。本項で、ニューラル機械翻訳で長大な請求項の翻訳精度が低下するメカニズムと、改善策を考察する。

第一の課題として、機械翻訳が一括で処理できる文長（トークン数）に上限があり、この上限が特許請求項の文長に対して不足していることが挙げられる。本調査分析でも、翻訳処理単位が文長の上限を超えた結果、翻訳文の出力が途絶したと見られるケースが散見された（⇒第 3.3.2 項）。また、長大な請求項では構成要素やステップなど適度な区切りで意図的改行やセミコロンが挿入されていることが多いが、このようにエンジンが文を分割できる状況でも、大規模な訳抜けが発生するケースが目立った（⇒第 3.3.3 項）。請求項の中盤以降で大規模な訳抜けが発生することが多いことから、エンジンの記憶領域の不足、もしくは前処理での文の分割が現状の記憶領域のサイズに合致していないことが原因である可能性がある。

加えて、ニューラル機械翻訳方式には、長文の翻訳精度を低下させる更なる事情もある。ニューラル機械翻訳では、入力文中の各用語とその周辺の語句との関係性をベクトル化して、それに最も近似のベクトル値をもつ訳語を学習データから採用しているが、エンジンが用語と周辺語句との関係性を把握できる範囲（距離）には限度がある。文長が極めて長く、密接に関連する語句（例えば、主語と述語）の間の距離が極度に遠くなると、エンジンは両者の関係を把握できず、適切なベクトル値を付せない。このため、誤訳や訳抜けが生じやすくなっている。長大な請求項の翻訳精度が低下する一因には、こうした事情も関与していると考えられる。

長大な請求項の大多数は、独立請求項である。独立請求項には発明の全体像が示されており、発明を理解する上で最も重要といえる。従属請求項は、概して短文であるため翻訳精度は高いが、従属請求項の内容は、引用する独立請求項の内容を踏まえなければ本質的には理解できない。したがって、翻訳精度の低い独立請求項を引用している従属請求項

は、それ自体が正しく訳されていても、実際の利用性は低い。つまり、請求項の機械翻訳の利用性は独立請求項の翻訳品質に依るところが大きく、長大な独立請求項の翻訳精度向上は、機械翻訳プラットフォームの最重要課題に挙げられる。

近年、ニューラル機械翻訳の新方式として、Transformer が採用されることが増えている。Transformer 方式は、従来のニューラル機械翻訳の弱点であった、距離の隔たった語間の関係性の把握に優れる。このため、特に長文の翻訳精度が大幅に改善されている。特許請求項が概して長大であることから、Transformer 方式は特許請求項の翻訳への適性が高いと考える。事実、Transformer 方式を採用した当財団の機械翻訳エンジンでは、機械翻訳プラットフォームで大規模な不備が生じていた多くの長文で、不備が解消又は軽減している（第 3.8 項で後述）。機械翻訳プラットフォームが Transformer 方式のエンジンを導入済みであるか否かは非公開であるが、仮に未導入である場合、Transformer 方式のエンジンを採用することで、長大な請求項の翻訳精度が大きく改善される可能性がある。

なお、機械翻訳システムは、エンジンによる翻訳処理の前後に、入力文のトークン化、翻訳可能な長さへの文分割、プレースホルダー化といった前処理と、それらを一文に戻して翻訳文を生成する後処理が必要である。第 3.3 項で述べたように、原稿の機械翻訳プラットフォームにはこうした点についての課題もあると見られる。長大な請求項の翻訳精度向上には、こうした翻訳の前後処理の改良も必要と考えられる。

3.7.5. 事例集

本調査分析で検出した、請求項の記載様式に係る不備事例を掲載する。

［事例 3.7.5-1：CN106659153B ｾｸｼｮﾝ A］

原　　文	:
	e)相对于总的组合物为 50g/kg 至 250g/kg 的固体缓冲剂，其中所述固体缓冲剂为硫酸铵或柠檬酸；和
	f)崩解剂，其中所述崩解剂为脲醛共聚物。
機械和訳	:
	e)全組成物に対して 50g/kg から 250g/kg までの固体緩衝剤(ここで、固体緩衝剤は硫酸アンモニウムまたはクエン酸である);および固体緩衝剤。
	f)崩壊剤が尿素尿素コポリマーである、崩壊剤。
校正結果	:
	e)全組成物に対して 50g/kg～250g/kg の固体緩衝剤であって,ここで,固体緩衝剤は硫酸アンモニウムまたはクエン酸である;および

	f)崩壊剤であって,ここで前記崩壊剤は尿素ホルムアルデヒドコポリマーである。
解　　説	第 3.7.2 項で分析した、発明の主題を文末に重複出力する傾向に該当する事例である。機械和訳は、最終文（「崩壊剤」が重複出力）だけでなく、2 文目でも「固体緩衝剤」が重複出力されている。2 文目は、セミコロンの後に「和（および)」があり、その後に意図的改行が挿入されているため、「和」までが一翻訳処理単位として扱われている。この状況で、文の主題を文末に重複出力したため、「および固体緩衝剤。」という不正確な翻訳文となっている。この行の末尾には句点が出力されており、翻訳処理がここで一旦完結している。このため「請求項の末尾で主題を繰り返す」という条件に合致したと考えられる。

3.8. サンプル 160 件における特徴的な不備の発生状況一覧

本章で言及してきた 160 件のサンプルにおける特徴的な不備のカウント結果について、報告書項番と対応させた一覧表を示す。なお、各サンプルにおける具体的な不備の内容は添付資料 02『サンプル 160 件における特徴的不備の検出記録』を参照されたい。

表 3.8-1.サンプル 160 件における特徴的な不備の発生数カウント結果一覧

不備の類型	事例	発生数	発生率
3.2. 用語の誤訳			
3.2.1 学習データ中に存在しない語の誤訳			
学習データに存在しない新語／マイナー語の誤訳	事例 3.2.1-1、3.2.1-3、3.2.1-4、3.2.4-2、3.2.4-4、3.2.4-5	44 件	27.5%
3.2.2 学習データ中に適訳語が存在する語の誤訳			
用語が正しい単位で認識されないことによる誤訳	事例 3.2.2-1	27 件	16.9%
訳語候補が複数存在することによる誤訳	事例 3.2.2-2、3.2.4-3	19 件	11.9%
3.2.3 係り受けの誤訳			
不適切な助詞による係り受け誤り	事例 3.2.3-1	21 件	13.1%
不適切な語順による係り受け誤り	事例 3.2.3-2	63 件	39.4%
3.2.4 事例集			
エンタメ系の誤訳	事例 3.2.4-1	0 件	0.0%
中国語の残存（未知語）	事例 3.2.4-4	13 件	8.1%
当て字の誤訳	事例 3.2.4-5	2 件	1.3%
3.3. 訳抜け			
3.3.1 用語レベルの訳抜け			
難訳語の訳抜け	事例 3.3.1-1	9 件	5.6%
本来訳出できる語の突発的な訳抜け	事例 3.3.1-2	47 件	29.4%
3.3.2 入力文字数超過による訳抜け			
長文の末尾で翻訳文が途絶する	事例 3.3.2-1、3.3.2-2	11 件	6.9%

3.3.3 セミコロン区切り文での訳抜け			
セミコロン区切り文での翻訳単位レベルの大規模な訳抜け	事例 3.3.3-1	53 件	33.1%
請求項の途中で原文にない句点が入る	事例 3.3.3-1	40 件	25.0%
3.3.4 事例集			
用語の羅列での訳抜け	事例 3.3.4-1	1	0.6%
3.4. 湧き出し			
3.4.1 一般的な湧き出し			
訳抜けを伴わない湧き出し	－	22 件	13.8%
訳抜けの補填的な湧き出し	事例 3.4.1-1、3.4.1-2、3.4.3-1、3.4.3-2	39 件	24.4%
3.4.2 不要な特許用語の湧き出し			
不要な特許用語の湧き出し	事例 3.4.2-1、3.4.2-2	8 件	5.0%
3.5. 数値の処理			
3.5.1 文末への数値出力			
訳抜け箇所の数字の羅列	事例 3.5.1-1、3.5.1-2	26 件	16.3%
3.5.2 漢数字の処理に係る不備			
漢数字から算用数字が派生	事例 3.5.2-1、3.5.2-2、3.5.5-5	12 件	7.5%
派生した数字が文脈に組み込まれる	事例 3.5.2-3、3.5.5-2、3.5.5-3	7 件	4.4%
「両」から算用数字 2 が派生	－	17 件	10.6%
3.5.3 漢数字を含む化合物名称における誤訳			
漢数字を含む化学物質名の誤訳	事例 3.5.3-1、3.5.5-2	7 件	4.4%

3.5.4 漢数字「一」の処理			
訳出すべき「一」の欠落 ／訳出すべきでない「一」の訳出	事例 3.5.4-1、3.5.5-3、3.5.5-6	7 件	4.4%
3.5.5 事例集			
請求項番号の複雑な羅列	事例 3.5.5-1	0 件	0.0%
「S311」等の誤訳	事例 3.5.5-4	0 件	0.0%
「一一対応」など「一一」の誤訳	事例 3.5.5-7	2 件	1.3%
3.6. 英数字記号の処理			
3.6.1 数式			
数式の反映ミス	事例 3.6.1-1、3.6.1-2	0 件	0.0%
3.6.2 化合物名			
英数字を含む化合物名の反映ミス	事例 3.6.2-1、3.6.2-2	0 件	0.0%
3.6.3 事例集			
中国語を含む数式の誤訳	事例 3.6.3-1	0 件	0.0%
長大な英数字文字列の反映ミス	事例 3.6.3-2	0 件	0.0%
3.7. 請求項の記載様式に係る不備			
3.7.1 請求項番号が本文の一部として扱われる不備			
請求項番号が本文の一部として扱われる不備	事例 3.7.1-1、3.7.1-2	13 件	8.1%
3.7.2 発明の主題を文末に出力する傾向			
主題を文末へ配置／再出力	事例 3.7.2-1、3.7.2-2、3.7.5-1	29 件	18.1%
主題を文末へ再出力した結果の誤訳	事例 3.7.2-2、3.7.5-1	7 件	4.4%
3.7.3 発明構成要素の訳ゆれ			
発明構成要素の訳ゆれ	事例 3.7.3-1	53 件	33.1%

3.9. 他種のニューラル機械翻訳エンジンとの翻訳結果比較

本章で不備の具体的事例として使用した各請求項について、他種のニューラル機械翻訳エンジンで翻訳を行い、同様の不備が生じるか等を調査した。本項にその結果をまとめる。

3.9.1. 比較に使用したエンジン

機械翻訳プラットフォームとの比較に使用したエンジンは下記 4 種である。

［エンジン①］

当財団の商用ニューラル機械翻訳サービス「Japio-AI 翻訳」。中日対応特許由来の対訳コーパス（学習データ量は非公開）を学習した特許文献専用の機械翻訳サービスであり、エンジンは Transformer 方式を採用している。中国語は自動判定を使用した。

［エンジン②］

当財団が研究用に構築したニューラル機械翻訳エンジン。エンジン①とほぼ共通の学習データ（ただし 15%ほど増量）を使用しているが、非 transformer 方式(RNN 方式で学習)である。本エンジンでは、入力文をできるだけ長い単位で翻訳した場合（つまり通常の場合）と、短く区切って翻訳した場合の二種類を作成し、より優れていると判定した方を採用する、一種のアンサンブル方式を採用している（ただし入力文が 300 文字を超える場合はエラー回避のため必ず短く区切って翻訳する）。中国語は簡体字を使用した。

［エンジン③］

ウェブ上で利用可能な自動翻訳サービス。Transformer 方式のニューラル機械翻訳エンジンを採用している。いくつかのオプションがあるが、本調査には特許翻訳オプション、中国語は簡体字を使用した。学習データには中日対応特許公報由来のコーパスが使用されていると推測される。

［エンジン④］

オンラインサービス企業がウェブ上で提供する自動翻訳サービス。ニューラル機械翻訳エンジンを採用しているが、Transformer 方式の採否は不明。学習データの詳細も非公開だが、特許文書に特化したサービスではなく、汎用的なものであるため、学習データも多種と推測される。中国語は簡体字を使用した。

下表に、各エンジンの各事例での不備の発生状況を一覧で示す。原則、各事例でトピックとなった不備が発生しているか、例えば事例 3.2.1-1 ならば「刷単行为（架空注文行動)」という用語が誤訳されているか否かを判定し、誤訳されていれば「×」、妥当に訳さ

れていれば「○」とした。なお、当該不備と別種の重大な不備が発生している場合は▲、当該不備が軽微になっている場合は△とした。なお、各エンジンによる各事例の翻訳結果と、その比較分析結果は、添付資料 03『他種エンジンによる機械和訳比較結果』を参照されたい。

表 3.9.1-1.各種ニューラル機械翻訳エンジンでの不備の発生状況一覧

事例	不備内容	ｴﾝｼﾞﾝ①	ｴﾝｼﾞﾝ②	ｴﾝｼﾞﾝ③	ｴﾝｼﾞﾝ④
3.2. 用語の誤訳					
3.2.1-1	「刷单行为」誤訳	×	×	×	×
3.2.1-2	「辅助口部防护装置」妥当訳 ※不備無し	○	○	○	×
3.2.1-3	「深度学习」誤訳	○	○	○	○
3.2.1-4	「松锤研磨」「紧锤研磨」誤訳	×	×	△	△
3.2.2-1	「风湿度」誤訳	○	○	×	○
3.2.2-2	「上网」誤訳	×	○	○	×
3.2.3-1	係り受け誤訳	○	○	○	○
3.2.3-2	係り受け誤訳	○	○	○	×
3.2.4-1	「打手」エンタメ系誤訳	▲	▲	▲	○
3.2.4-2	「降焦烟用接装纸」誤訳、訳ゆれ	×	×	×	▲
3.2.4-3	「研磨」誤訳	×	○	○	○
3.2.4-4	「螺纹举升机构」一部未知語	○	○	○	▲
3.2.4-5	当て字の誤訳	×	×	○	○
3.3. 訳抜け					
3.3.1-1	「当期销售器」訳抜け	○／×	×／×	×／×	○／×
3.3.1-2	「细胞表面标记」突発的訳抜け	○	×	○	▲
3.3.2-1	長文末尾での翻訳文途絶①	○	▲	▲	▲
3.3.2-2	長文末尾での翻訳文途絶②	○	▲	×	▲
3.3.3-1	セミコロン区切り単位の訳抜け	○	○	△	○
3.3.4-1	鉱物名称羅列での訳抜け	○	○	○	△
3.4. 湧き出し					
3.4.1-1	訳抜けの補填的な湧き出し①	○	▲	○	▲
3.4.1-2	訳抜けの補填的な湧き出し②	○	○	○	▲
3.4.2-1	不要な特許用語の湧き出し①	○	○	×	○
3.4.2-2	不要な特許用語の湧き出し②	○	○	○	○
3.4.3-1	菌名羅列での訳抜けと湧き出し	×	×	△	▲
3.4.3-2	意味不明な文言の湧き出し	▲	▲	▲	▲

3.5. 数値の処理					
3.5.1-1	訳抜け箇所の数字の文末羅列①	○	○	○	○
3.5.1-2	訳抜け箇所の数字の文末羅列②	○	○	○	▲
3.5.2-1	漢数字から算用数字が派生①	○	○	○	○
3.5.2-2	漢数字から算用数字が派生②	○	○	○	○
3.5.2-3	「両」から算用数字 2 が派生	○	○	○	○
3.5.3-1	漢数字を含む化学物質名の誤訳	○	○	○	○
3.5.4-1	訳出すべき「一」の欠落	○	○	○	▲
3.5.5-1	請求項番号の複雑な羅列	○	○	○	○
3.5.5-2	漢数字から算用数字が派生③	○	○	○	▲
3.5.5-3	漢数字から算用数字が派生④	○	○	○	○
3.5.5-4	「S311」等の誤訳	○	○	○	○
3.5.5-5	「“十”形」の不適訳	○	○	○	×
3.5.5-6	「一」の訳出の要否判断 ※不備なし	○	○	○	○
3.5.5-7	「一一対応」の誤訳	○	○	○	○
3.6. 英数字記号の処理					
3.6.1-1	数式の反映①※不備なし	○	○	○	○
3.6.1-2	数式の反映②※不備なし	○	○	○	○
3.6.2-1	英数字を含む化合物名①	○	△	○	×
3.6.2-2	英数字を含む化合物名②	×	×	×	×
3.6.3-1	中国語を含む数式	×	○	○	×
3.6.3-2	英数字文字列の長大な羅列 ※不備なし	○	○	○	×
3.7. 請求項の記載様式に係る不備					
3.7.1-1	請求項番号の誤解釈①	○	○	○	○
3.7.1-2	請求項番号の誤解釈②	○	○	○	▲
3.7.2-1	主題を文末へ再出力①	×	○	×	○
3.7.2-2	主題を文末へ再出力②	▲	▲	○	○
3.7.3-1	発明構成要素「推進器」の訳ゆれ	×	△	△	△
3.7.5-1	文末への不要な文言の出力	○	○	○	○

※表中「※不備無し」とした事例は、機械翻訳プラットフォームにおいても不備が発生していない。

全事例における各エンジンの不備の発生状況を集計した結果を下表に示す。

表 3.9.1-2.各種ニューラル機械翻訳エンジンでの不備の発生状況の集計結果

	○	△	▲	×
エンジン①	37	0	3	12
エンジン②	35	2	6	9
エンジン③	36	4	3	9
エンジン④	26	3	13	10

上表「×」欄のとおり、機械翻訳プラットフォームで発生した不備の多くは、他種エンジンでは発生していない。ただし、ニューラル機械翻訳の不備には突発性が強いものも多く、今回用いた事例では不備が発生しなかったエンジンも、他の文で同種の不備が発生しないとは限らない。さらに、そうした文が機械翻訳プラットフォームでは適切に訳される可能性も十分にある。今回の調査方法は、機械翻訳プラットフォームに不利な条件であり、調査結果は各エンジンとの性能比較には適さない。

本調査は、他種エンジンの翻訳結果と比較することで、機械翻訳プラットフォームに特徴的な不備のうち、他のエンジンでは解決されており、したがって機械翻訳プラットフォームでも解決可能と見なせる課題を見極めるためのものである。以下、その見地から上記結果を不備の類型別に分析していく。

3.9.2. 用語の誤訳

機械翻訳プラットフォームで誤訳されたマイナーな技術用語、具体的には「刷単行为」(事例 3.2.1-1）や「松锤研磨／紧锤研磨」(事例 3.2.1-4)、「降焦烟用接装纸」(事例 3.2.4-2）に関しては、他のエンジンでも同様に誤訳された。これらの用語はいずれも中日対応特許での使用例がなく[69]、学習データ自体が流通していないと考えられる。ニューラル機械翻訳では、学習データに存在しない語は翻訳できない。

一方、「深度学习」に関しては、各エンジンで「ディープラーニング」又は「深層学習」と適切に翻訳された。当財団の中日対応特許コーパスにも大量の訳例が存在しており、学習データがすでに流通している状況である。このことから、機械翻訳プラットフォームは、学習データのアップデートが他のエンジンより遅れていると推測される。第 3.2.1 項では、こうした使用頻度の高い新語は新規の中日対応特許由来のコーパスを定期的に学習させることで自動的にカバーされると述べたが、本調査の結果はこれを裏付けている。

[69] 第 3.2.1 項の脚注 37 を参照。

複数の解釈や複数の有力訳語が存在する「风湿度」(事例 3.2.2-1) や「上网」(事例 3.2.2-2)、「研磨」(事例 3.2.4-3)、そして採用された語順や助詞などによって発生する係り受け誤り（事例 3.2.3-1、3.2.3-2）に関しても、エンジンによっては正しく訳出できていた。ただし、これらの不備は突発性が強く、今回の結果のみで優劣は測れない。

3.9.3. 訳抜け

難訳語の訳抜け事例として取り上げた「当期销售器」(事例 3.3.1-1) に関しては、エンジン①で妥当訳に訳出された。一方、同じ事例中の別の難訳語「预售销售器」に関しては、エンジン④で妥当な訳語が採用されている。ただし両語とも他のエンジンでは誤訳もしくは訳抜けが生じており、機械翻訳プラットフォームと状況は大きくは違わない。この結果から、両語は学習データ、すなわち中日対応特許中に訳例が乏しいマイナーな用語と見られる。

機械翻訳プラットフォームでは、特徴的な訳抜けの類型として、長文における翻訳の途絶（事例 3.3.2-1、3.3.2-2）や部分的な訳抜け（事例 3.3.3-1、3.3.4-1）も見られた。第 3.7.4 項では、こうした長文の翻訳精度の向上に Transformer 方式の採用が有効である可能性を述べたが、Transformer 方式を採用しているエンジン①では、これら全ての事例で長文が正確に処理されていた。一方、同じく Transformer 方式のエンジン③では、事例 3.3.3-1 や 3.3.4-1 など適切に翻訳された事例もある反面、事例 3.3.2-1 や 3.3.2-2、そして事例 3.4.3-2 や 3.2.4-2 の請求項 1 など長大な請求項では翻訳文の途絶が頻繁に発生した[70]。

これらの結果は、Transformer 方式が長文の翻訳精度の改善に有効であることと、ただし Transformer 方式の採用のみで全ての長文の翻訳精度が改善されるとは限らないことの双方を示している。第 3.7.4 項で述べたとおり、長文の翻訳精度には翻訳エンジンの方式だけでなく、適切な前処理・後処理といった周辺処理も重要であることがわかる。

なお、非 Transformer 方式であるエンジン②でも長文での訳抜けは発生していないが、これは、本エンジンが 300 文字を超える長文を強制的に細かく区切って個別に翻訳しているためである。この処理によって、長文を翻訳処理することで多発しがちな訳抜け等のエラーは回避されている。その反面、事例 3.3.2-1、3.3.2-2、3.2.4-2 などの長文では、節ごとに終止形や体言止めが用いられて不自然な文になるという別の問題が生じている。また、この手法には、節を跨いだ係り受け関係が不正確になるという短所もある[71]。

---

[70]長文での翻訳文の途絶は、事例 3.2.4-2 の請求項 1 のようにセミコロンで適度な文長に区切られている請求項でも見られた。

[71] 例えば事例 3.3.2-2 冒頭部の「所述相位解调装置包括窄线宽激光器、第一光耦合器、光调制器、光放大

3.9.4. 湧き出し

訳抜けの補填的な湧き出し（事例 3.4.1-1、3.4.1-2）に関しては、使用した各事例では各エンジンとも訳抜け自体が（長文における翻訳文の途絶を除けば）ほとんど発生していないため、その補填としての湧き出しもほとんど検出されなかった。

不要な特許文献用語の湧き出しに関しても、機械翻訳プラットフォームで不備が生じた事例（事例 3.4.2-1、3.4.2-2）のうち、他のエンジンで不備が発生したのは一例（エンジン③、事例 3.4.2-1）のみであった。ただし、エンジン③とエンジン④では別カテゴリの事例（事例 3.3.3-1、3.3.1-2）中で同種の不備が発生しており、このタイプの湧き出しエラーが突発的に発生する性質であることがうかがえる。第 3.4.2 項で述べたとおり、この不備は特許文献由来の大量の学習データを学習させたことによる副作用の一種と考えられ、今回使用した事例で湧き出しが発生しなかった他のエンジンにおいても、この種の不備が完全に解決されているとは言い切れない。

3.9.5. 数値の処理

数値の処理に係る不備に関しては、各エンジン、全ての事例において、機械翻訳プラットフォームと同種の不備は発生しなかった。機械翻訳プラットフォームでは、数値関連の不備はそれぞれ特定の条件で頻繁に発生しており、他の多くの不備と異なり法則性が高い。こうした条件に該当する各事例で不備が発生しなかったことから、機械翻訳プラットフォームで見られた数値関連の各種不備は、他種のエンジンでは一切発生しないと考えられる。本調査結果は、機械翻訳プラットフォームにおける数値処理関連の不備が、学習データの内容ではなく、システムの処理方式自体に起因している可能性が高いという第 3.5 項の考察を裏付けるものである。

器，环形器、第二光耦合器、传感光纤、脉冲信号发生器、光电探测器和数据采集器,」の「包括（含む）」は、後方に列挙された 9 つの構成要素に係っている。このため「前記位相復調装置は狭線幅レーザ,第一光カプラ,光変調器,光増幅器,サーキュレータ,第二光カプラ,センシングファイバ,パルス信号発生器,光検出器及びデータ収集器を含み,」と訳すのが正しいが、エンジン②の翻訳文は「前記位相復調装置は、狭線幅レーザと、第 1 光カプラと、光変調器と、光増幅器と、を備え、サーキュレータ、第二光カプラ、センシング光ファイバ、パルス信号発生器、光検出器及びデータ収集器,」と、前半 4 つの構成要素にしか係っていない。これは、この節がこの位置で分割されて別個に翻訳されたことを示している。その結果、節の後半は動詞不在の名詞の羅列となり、体言止めで訳されたことで、文法的に成立しない翻訳文となってしまっている。

3.9.6. 英数字記号の処理

英数字記号の処理に関しては、機械翻訳プラットフォーム自体、ほとんど不備は発生しなかったが、他のエンジンも（エンジン④を除き）同様の状況であった。ニューラル機械翻訳において、入力文中の英数字文字列は、ほぼ不備の原因とはならないことが改めて確認された。

3.9.7. 請求項の記載様式に係る不備

請求項の記載様式に係る不備に関しては、請求項番号が本文の一部として扱われる重大な不備（事例 3.7.1-1、3.7.1-2）は他のエンジンでは発生しなかった。ただし、この不備も（請求項番号を本文と切り離して処理する等の特殊な前後処理を施さない限りは）突発性を有すると考えられ、必ずしも他のエンジンで一切発生しないとは言い切れない。

一方、発明の主題とみなした語を文末に出力する傾向（事例 3.7.2-1、3.7.2-2）に関しては、各エンジンとも多かれ少なかれ同じ傾向が見られた。これは、日本の記載様式で書かれた特許請求項に由来するコーパスを学習データに用いている以上、必然的に生じる傾向であるといえる。また、発明構成要素の訳ゆれ（事例 3.7.3-1）に関しても、各エンジンとも機械翻訳プラットフォームと同様に発生しており、現状においては、ニューラル機械翻訳の宿命的な事象であることがわかる。

3.9.8. まとめ

他エンジンとの比較結果を総合的に判断すると、機械翻訳プラットフォームは、学習データの量と新しさ、長文の翻訳精度、そして数値処理に関して、改善の余地があると結論できる。学習データの量・新しさに対しては中日対応特許由来の学習データの定期的なアップデート、長文の翻訳精度に対しては、仮に未採用であれば Transformer 方式の採用と適切な前後処理、そして数値処理に対しても改善が必要と考えられる。一方、マイナーな技術用語の誤訳や発明構成要素の訳ゆれ、発明主題の文末出力の傾向は他の各エンジンでも同様に検出されており、現時点では解決困難な、ニューラル機械翻訳共通の課題であることが再確認された。

4. 本調査のまとめ

本調査により把握した、機械翻訳プラットフォームにおける中国特許請求項の機械和訳の翻訳品質及び不備の傾向と、さらなる精度向上のための提言を本章にまとめる。

4.1. 機械翻訳プラットフォームの現状の翻訳品質について

本調査では、中国特許公報 10 万件の全請求項 949,882 項に対する校正作業の記録に基づき、現状の機械翻訳プラットフォームにおける機械和訳の不備の状況を調査分析した。

第 2 章に示した統計的分析では、校正者によって何らかの修正が加えられた請求項は全体の 86%を超えたものの、発明の内容理解を大きく阻害するレベルの「重大な不備」の発生率[72]は 21.4%に抑えられていることがわかった。ただし、独立請求項と従属請求項に区別した集計では、独立請求項において重大な不備の発生率が 67.1%と非常に高いことも判明した。従属請求項の重大不備発生率は 13.4%と低かったが、従属請求項は引用した独立請求項の内容を踏まえて理解すべきものであり、したがって機械翻訳プラットフォームの実質的な翻訳精度は独立請求項を基準に判断するのが妥当である。つまり、現状の機械翻訳プラットフォームにおける最大の課題は、67.1%と高頻度で発生している独立請求項の重大不備を減少させることにある。

独立請求項で重大不備が多発する最大の要因は、その極端に長大な文長にあると考えられる。本調査の対象の 10 万件においても、従属請求項の平均文長が 88.7 文字であったのに対し、独立請求項の平均文長は 400 文字を超えていた（420.4 文字）。特許請求項は一文で記述するのが通例であるため、発明の全体像が詳細に記載される独立請求項は、一般的な文に比べて極端な長文となる傾向が強い。

機械翻訳は一文単位の処理が基本であり、その翻訳精度は原則として文長が長くなるほど低下していく。事実、本調査の統計的分析でも、請求項の文長と重大不備の発生率との間に明瞭な相関性が確認された（⇒第 2.5 項）。具体的には、「重大な誤訳」が文長に比例してコンスタントに増加していくことに加え、文長が 150 文字を超えると、それまで少数であった「重大な訳抜け」と「重大な湧き出し」も急増する傾向が見られた。この傾向は基本的には独立請求項、従属請求項の双方に共通する。従属請求項の場合、独立請求項よりも不備の発生率が若干低かったが、これは従属請求項に翻訳難易度の低い定型の文パターンが用いられることが多いためと考えられる（⇒第 2.5.2 項）。150 文字を超えた各文長

[72] 重大な不備の発生率は、事業後期分の集計値に基づいた。

範囲での重大不備の発生率は独立請求項より若干少ない程度で大差はなく、長大な請求項では種別によらず重大不備が多発している状況といえる。

なお、長大な請求項では、意図的改行やセミコロンが多用される。機械翻訳プラットフォームでは、実際の機械和訳の状況から、他の多くの機械翻訳システムと同様、これらを文の分割位置として利用していると推測される。つまり、長大な請求項も意図的改行／セミコロンで適宜分割され、実際の機械翻訳処理の単位は適度な文長に抑えられていると考えられる。一文を分割して翻訳する手法は、分割された各部分間の係り受け関係が失われるデメリットがあるが、請求項の場合、例えばステップや構成要素ごとに意図的改行・セミコロンで区切られ、内容的にも文法上も各区切り単位で文として独立していることが多い。このため、分割して翻訳し、単純に連結する方法でも、翻訳文が破綻なく成立することが多いと予測された。

事実、統計的分析でも、同程度の長文であれば、適度に意図的改行やセミコロンが挿入されているもののほうが重大不備の発生率は低かった（⇒第 2.5.3 項）。だが、これを踏まえても長大な請求項における不備の発生率は依然として高く、「重大な訳抜け」や「重大な湧き出し」が顕著に増加する傾向も、意図的改行やセミコロンの多寡にかかわらず見られた。機械翻訳プラットフォームではセミコロン等で必ずしも文分割を行わない可能性も考えられるが、多数の不備事例の状況から総合的に判断すると、長文の処理において文分割のみでは解決されない課題がこれらの結果に反映されている、と考えられ、長文の翻訳品質において大きな改善の余地がある。

統計的分析ではさらに、技術分野（IPC セクション）別、出願人の国籍別、法人／個人の別などの観点でも不備の発生頻度を調査した。その結果、各観点とも文長差に準拠した不備発生率の差異が生じており（⇒第 2.6～2.8 項）、請求項の文長が機械翻訳プラットフォームの翻訳精度を決定する最大のファクターであることが改めて示された。

4.2. 機械翻訳プラットフォームの特徴的な不備と対処策について

本調査ではさらに、現行の機械翻訳プラットフォームに見られる特徴的な不備を収集し、類型別に分析した（⇒第 3 章）。その結果、中国特許請求項において重大な不備をもたらす主な要因、つまり最優先の課題として、以下の四点が挙げられた。

- 学習データが最新の技術用語に対応できていない
- 前後処理を含めた長文の処理精度が不十分である
- 数字（漢数字を含む）の処理方式に改善の余地がある
- 請求項冒頭の請求項番号が本文に組み込まれて誤解釈される

以下、上記各点について、具体的な状況と、採り得る改善策を提言する。

4.2.1. 学習データが最新の技術用語に対応できていない

特許文献用のニューラル翻訳システムでは、対応特許由来のコーパスが学習データに用いられることが多い。機械翻訳プラットフォームも、学習データの内容は非公開ながら、主に対応特許コーパスを使用している可能性が高い。

ニューラル機械翻訳では、学習データ中に存在しない語彙は正しく翻訳することができない。請求項には非常にマイナーな技術用語や出願人の造語が使われることもあり、対応特許由来の学習データでこれら全ての語彙をカバーすることは不可能である。このため、ある程度の誤訳、不適訳の発生は事実上避けられない。

しかしながら、本調査の結果、機械翻訳プラットフォームでは「ディープラーニング（深度学习）」や「ビッグデータ（大数据）」、「充電スタンド（充电桩）」など、近年になり一般的に使われ出したメジャーな技術用語もほぼ常時誤訳されている[73]ことが明らかとなった（⇒第 3.2.1 項）。これらの語句は中日対応特許でも既に多用されており、他種のニューラル機械翻訳エンジンでは概ね正しく翻訳される（⇒第 3.9.2 項）。

こうした結果から、機械翻訳プラットフォームの機械翻訳エンジンは、他のエンジンに比べてアップデートが遅れている可能性が高い。特許文献は最先端の技術が記載される文書であることから、定期的に最新の対応特許由来のコーパスを追加学習させ、語彙を増強した機械翻訳エンジンを用いることが有効であると考える。学習データの増強は、マイナーな技術用語のカバー範囲を広げることにもつながる。

---

[73] 「ディープラーニング（深度学习）」は 184 項中 183 項（99%）、「ビッグデータ（大数据）」は 449 項中 408 項（91%）、「充電スタンド（充电桩）」は 141 項中 127 項（90%）と大半が誤訳されている。正しく訳されているものは、用語を分解して個別に訳した結果、偶然正しい訳語となったと考えられる。

4.2.2. 前後処理を含めた長文の処理精度が不十分である

第 4.1 項で述べたとおり、機械翻訳プラットフォームでは、特に文長 150 文字を超える請求項において、内容の大規模な欠落（訳抜け）や、訳抜けした箇所への文中の別の語句の重複出力（湧き出しの一種）といった重大な不備が多発する傾向が顕著であった。この傾向は、独立／従属請求項、技術分野、出願人の国籍、意図的改行／セミコロンの多寡などにかかわらず一律に見られ、現行の機械翻訳プラットフォームが、特許請求項の「一文で記載されるため極端に長大になる傾向が強い」という特徴に十分に対応できていないことを示している。

長文の翻訳精度改善の一手段として、新方式のニューラル機械翻訳である Transformer 方式の採用が推奨される。この方式は、従来のニューラル機械翻訳の弱点であった、距離の隔たった語間の関係性の把握に優れる。このため、特に長文の翻訳精度改善に有効であり、特許請求項の翻訳用途への適性が高いと考える。事実、Transformer 方式を採用した当財団の機械翻訳エンジンでは、機械翻訳プラットフォームで大規模な不備が生じた多くの長文で、不備が解消又は軽減している（⇒第 3.9 項）。

ただし、長文の翻訳精度改善には、翻訳の前後処理の精度も等しく重要である。機械翻訳プラットフォームでは、長大な請求項は意図的改行やセミコロンで適度な文長に区切って翻訳されていると推測されるが、こうして分割して翻訳された請求項においても、大規模な訳抜けや湧き出し等の不備は依然多発しており、顕著な改善は見られない（⇒第 2.5.3 項、第 3.3.3 項）。類型ごとの文長別発生率（⇒図 2.5.1-1）を見ても文長 150 文字以下の短い請求項では重大な訳抜けや湧き出しがほとんど発生していないことから、長文を区切って順番に機械翻訳エンジンに供する前処理、又は順次翻訳された各文を一文に連結して出力する後処理においても改善の余地がある可能性が高い。

第 3.9 項に示した他種エンジンとの比較でも、Transformer 方式を採用したエンジンの中にも機械翻訳プラットフォームと同様の不備（長文における翻訳文の途絶）が頻繁に発生しているものがあり、長文の翻訳精度向上は Transformer 方式の採用のみで達成されるとは限らないことがわかる。機械翻訳プラットフォームにおいて長大な請求項の精度向上を果たすには、Transformer 方式の採用と、長文に対する文分割・再結合などの前後処理の精度向上の双方が等しく重要であると考える。

4.2.3. 数字（漢数字を含む）の処理方式に改善の余地がある

機械翻訳プラットフォームでは、数字・漢数字に関連する特徴的な不備が非常に多く見られた（⇒第 3.5 項）。なかでも内容理解への悪影響が特に大きなものとして、原文中の漢数字が算用数字に置換された結果、漢数字を含む化合物名称などの技術用語が正しく扱われず、学習データ中に訳例が存在するにもかかわらず誤訳されてしまう不備や、漢数字から派生した算用数字が原文中の他の算用数字と置き換わって不正な内容になる不備などが挙げられる。

中日対応特許コーパス（すなわち学習データ）には、原文は漢数字、翻訳文では算用数字が用いられる対訳文も存在するが、他種エンジンでは同様の不備は見られない。したがって学習データが原因とは考えにくく、機械翻訳プラットフォームに搭載されたエンジンの数字・漢数字処理に起因した不備である可能性が高い。

特許請求項において数値は重要な情報である。加えて、漢数字は中国語では化合物名称をはじめ様々な技術用語に含まれており、数値関連の不備の影響範囲はきわめて広い。これらの不備が機械翻訳プラットフォームのエンジンに起因するものであるならば、その更新が必要となる。学習データのアップデート、長文の処理精度向上とともに、最優先で対処すべき課題であろう。

4.2.4. 請求項冒頭の請求項番号が本文に組み込まれて誤解釈される

第 3 章で「請求項の記載様式に係る不備」として挙げた「請求項番号が本文の一部として扱われる不備（⇒第 3.7.1 項）」も、請求項 1 では全件の 4.5%で発生するなど発生頻度が高く、かつ請求項番号が誤った解釈で不必要な文言を伴って本文に組み込まれることで翻訳文全体の翻訳精度に大きな悪影響を及ぼす点において、優先的に対処が必要な課題である。

この不備は長大な請求項に偏って発生していることから、第 4.2.2 項で述べた長文の処理精度という課題と関連している可能性が高い。かつ、数値が無関係な文言を伴って翻訳文に組み込まれるという点において、第 4.2.3 項で述べた数値処理関連の不備にも該当する。このため、これらへの対処により改善される可能性もある。または、第 3.7.1 項で述べたように、文頭の「数字＋ピリオド」は前処理であらかじめ本文と切り離し、機械翻訳処理を迂回させて、後処理で翻訳文に戻すという手法も有効であろう。

4.3. 現状では改善困難な不備

本調査で検出された不備の類型の中には、他種のエンジンでも等しく発生し、現状のニューラル機械翻訳技術では改善が難しいと考えられるものも存在した。以下、主要な二種について述べる。

4.3.1. マイナーな技術用語や出願人による造語の誤訳

「学習データ中に存在しない誤訳（⇒第 3.2.1 項）」の多くを占めたマイナーな技術用語や出願人の造語は、中日対応特許にも使用例が存在しない場合が多い。こうした語を正しく翻訳させるためには、対訳データを自ら作成して学習に供する必要がある。だが、これを行うには、まずは（本調査と同様に）機械和訳を解析して誤訳されている語句を収集する必要があり、さらに対訳データを人手翻訳で作成しなければならない。これには多大なコストを要し、かつ、こうした語は無限に発生するため際限のない対応となる。

マイナーな用語や出願人の造語はごく少数の案件のみで用いられる傾向にあるため、対応しても全体的な翻訳精度の改善効果は小さい。現状ではこうした語彙を網羅的にカバーできる決定的な打開策はなく、コストパフォーマンスの観点からの割り切りが必要であろう。

4.3.2. 同一の技術用語の訳ゆれ

特許請求項では、同一の技術用語が発明の構成要素として繰り返し言及される傾向が強い。同一案件中の多数の請求項に共通の技術用語が用いられることも通常である。

一方、ニューラル機械翻訳は、各用語の訳語を学習データに基づいて都度決定している。このため、同じ用語であっても分野や文脈に応じて自動的に訳し分けることができるが、その反面、同一案件や同一請求項において同じ語が異なる訳語に「訳ゆれ」する可能性も生じる。事実、本調査では、このような発明構成要素の訳ゆれが非常に多く検出された（⇒第 3.7.3 項）。

現状、こうした訳ゆれはニューラル機械翻訳の方法自体が内包する宿命的なものといえ、即効性のある対処策は見当たらない。ただし、ニューラル機械翻訳分野では現在、この課題への対処策として、前後の文の文脈を考慮して（つまり、先行文で用いた訳語情報を保存するなどして）訳語の統一を図る手法の研究が進められている。将来的には、請求項単位や案件単位で訳語情報を共有することで、訳ゆれの解消が進むものと期待される。

4.4. IPC セクション別／出願人の国コード別の翻訳精度の違いについて

本調査で行った IPC セクション別の統計的分析の結果、本調査対象においては、E セクション、D セクション、G セクションが他のセクションに比べて重大な不備の発生率が高いことが判明した（⇒第 2.6 項）。

これら 3 セクションは、請求項の平均文長が他より顕著に長く、このことが高い不備発生率の原因であると見られる。本調査の対象案件は、各セクションとも全案件の「特許請求の範囲」が平均 1,300 文字となるように選定したが（⇒第 1.3 項）、各案件の「特許請求の範囲」においてどのような請求項の構成となっているかは考慮していない。このため、これら 3 セクションは、中国特許公報全体においても他のセクションより請求項が長大となる傾向を有し、これにより機械和訳の品質が他より低くなっている可能性がある。その場合、第 4.2.2 項で述べたような長文の処理精度向上策により、他セクションとの品質差が縮まる可能性が高い。

また、E セクションと D セクションに関しては、主要な学習データである中日対応特許の件数自体が他のセクションより著しく少なく、学習データの充実度に差がある可能性も考えられる。第 2.6.1 項の表 2.6.1-3 に示したとおり、E、D セクションの構成比はそれぞれ 0.7%、0.9%と極端に低い。このため、長文の処理精度向上のみでは、他のセクションと同等の翻訳精度には至らないかもしれない。

だが、両セクションの学習データ量を他と同水準に増量するためには大規模な人手翻訳作業を行わなければならず、多大なコストが必要となる。さらに言えば、E セクション、D セクションは中国公開公報の発行件数自体も他のセクションに比べて非常に低いことがわかっており（4.8%、1.4%。第 2.6.1 項の表 2.6.1-4 参照）、マイナーな技術分野の学習データを極端に増強した結果、それが他の技術分野の翻訳に過度に反映され、全体の翻訳精度に悪影響を及ぼすというリスクも懸念される。

上記した中国公開公報の E、D セクションの構成比は、機械翻訳プラットフォームで実際に翻訳対象となる中国文献の構成比と見なせる。この観点からは、既存の中日対応特許から作成されたコーパスをそのまま学習させることが、翻訳対象の文献集合の構成比に則したバランスの良い学習となると考える。

一方、出願人の国コード別の統計的分析では、中国出願人から中国への出願（CN 出願）の重大不備発生率が諸外国から中国への出願（外国出願）に比べて高かった。これも主として双方の平均文長の差によるものと結論される。CN 出願は、（理由は不明ながら）

一案件あたりの従属請求項数が少なく、かつ独立請求項、従属請求項とも一請求項あたりの平均文長が外国出願に比べて長大となる傾向が見られた。（⇒第 2.7 項）

また、外国出願のうち日本からの出願（JP 出願）は、CN 出願や他の外国出願に比べて重大不備の発生率が低いことも判明した。ただし、これは機械翻訳プラットフォームの主要な学習データと想定される中日対応特許コーパスに当該出願そのものが含まれる確率が高いためと考えられる。

中日対応特許としてコーパス化される案件は、ファミリー中に中国特許出願と日本特許出願の両者が存在するものである。本調査でいう CN 出願、（JP 出願を含む）外国出願とも全て中国特許出願であり、これらのうち日本のファミリー文献を有する確率は、出願人が日本在住である JP 出願が最も高いことは自明である。したがって、JP 出願は、中日対応特許コーパス中に自身が収録されている確率が他より高くなる。そして、入力文が学習データと完全一致していれば、翻訳精度において圧倒的に有利であることは言うまでもない。

よって、中日対応特許コーパスを学習データに用いる以上、JP 出願が他の出願に比べて重大な不備が少なくなることは必然である。学習データと完全一致する確率が他より高いことによるバイアスであり、CN 出願や JP 以外の諸外国出願の学習データが不足していることを意味しない。

同じ国からの出願であっても発明の内容はそれぞれ異なる。このため、「学習データ中に自身が収録される」ケースを度外視すれば、学習データにおける出願人の国籍の偏りが各国出願の翻訳品質に顕著な有利不利を生じるとは考えにくい。学習データ中の国籍のバランスを過度に懸念する必要はなく、入手できる中日対応特許コーパスを（国籍の比率を問わず）最大限に学習させるのが最善と考える。

添 付 資 料

「重大な誤訳」「重大な訳抜け」「重大な湧き出し」の判定基準

本事業では、機械和訳に含まれる不備のうち、原文の内容（特に技術内容）の正しい理解を大きく妨げるレベルの不備を「重大な不備」と見なし、校正者がその有無を判定した。なお、重大不備の有無は「重大な誤訳」「重大な訳抜け」「重大な湧き出し」の三類型別に行った。本資料にて、重大な不備に該当するものと、しないものを例示する。

１．重大な誤訳と、そうでないものの実例

ここで言う「誤訳」は、原文中の用語・表現が、翻訳されてはいるが適切な訳語にはなっていないことを指す。訳語自体が出力されていないケースや、原文中に存在しない余計な文言が出力されているケースは、それぞれ「訳抜け」「湧き出し」として扱う。

まずは「重大な誤訳」と判定すべき実例を示す。

［重大と判定すべき誤訳例］

原　文	2.根据权利要求 1 所述的自移动设备，其特征在于，所述自移动设备直线移动时，所述控制模块控制所述履带的接地长度大于或等于 300mm。
機械和訳	2.請求項 1 に記載のモバイル機器であって,その特徴は以下のとおりである:前記モバイル機器が直線移動する時に,前記制御モジュールが前記クローラの接地長を 300mm 以上に制御する。
校正結果	2.請求項 1 に記載の自動移動機器であって,その特徴は以下のとおりである:前記自動移動機器が直線移動する時に,前記制御モジュールが前記クローラの接地長さを 300mm 以上に制御する。
解　説	この発明は具体的には自走式の芝刈り機に関する。よって原文の「自移动设备」は「自動移動機器」等と訳すべきものである。校正前の機械和訳の訳語「モバイル機器」は、日本語では「持ち運びができる通信用機器」として使われる語であり、大きな誤解を生じかねない。この語は発明の主題でもあり、正しい内容理解を阻害する「重大な誤訳」と判定するのが妥当である。

上例での校正箇所はほぼ「自移动设备」の訳語の修正のみに限られ、物理的な校正量は少ない。だが、重大な不備であるか否かの判定は物理的な校正量ではなく「正しい内容理解への悪影響の多寡」という観点で行うため、上例のように小規模であっても重大と判定すべきケースもあり、逆に校正量が多くても重大と判定すべきでないケースもある。

続いて、重大と判定すべきでない誤訳例を示す。

［重大と判定すべきでない誤訳例：CN107723827B］

原　文	5.根据权利要求 1 所述的一种吸湿速干原液着色纤维的制备方法，其特征在于，步骤(2)中，所述色浆在线注入量为聚酯低聚物质量的 3～20wt%。
機械和訳	5.請求項 1 に記載の吸湿速度乾燥原液の着色繊維の製造方法であって,その特徴は以下のとおりである:ステップ(2)において,前記色ペーストのオンライン注入量はポリエステルオリゴマー物質の量の 3～20wt%である。
校正結果	5. 請求項 1 に記載の吸湿性速乾性原液着色繊維の製造方法であって,その特徴は以下のとおりである:ステップ(2)において,前記カラーペーストのオンライン注入量はポリエステルオリゴマーの量の 3～20wt%である。
解　説	本例の校正前の機械和訳には 3 箇所の誤訳がある。特に冒頭の「吸湿速干原液着色纤维」は、「吸湿性と速乾性を有し、原液で着色された繊維」を指しているが、機械和訳ではこの語が「吸湿速度乾燥原液の着色繊維」と訳され、各語の指す意味やそれぞれの関係性が不詳である。明らかな誤訳であり、校正は必須であるが、ただし発明の内容自体は類推可能であり、内容理解を致命的に阻害してはいないため、重大と判定するほどではない。

本例は複数の箇所に校正がなされており、冒頭部のような明らかな誤訳も含まれる。ただし、各誤訳とも読み手の理解を著しくミスリードするほどのものではなく、内容理解上致命的ではないため、「重大な誤訳」と判定するほどではない。

とはいえ、本例の誤訳箇所（特に「吸湿速度乾燥原液の着色繊維」）は、日本語として非常に低品質であり、理解しにくいことは確かである。重大か否かの判定は微妙で、本例はその境界線付近にあるといえる。換言すれば、本例よりも軽度の誤訳であれば、重大とは判定しない。たとえば本例の「色ペースト」や「ポリエステルオリゴマー物質の量」の箇所の誤訳は、「重大」とする必要のないレベルの誤訳である。

2．重大な訳抜けと、そうでないものの実例

「訳抜け」とは、原文中の要素が翻訳されず、翻訳文中から欠落することを指す。こちらも、重大と判定すべきなのは「正確な内容理解に大きな悪影響を及ぼす訳抜け」であり、全ての訳抜けをカウントする必要はない。以下、重大な訳抜けの実例を示す。

［重大とみなすべき訳抜け：CN106629937B］

原　文	25.按照权利要求 17 所述的工艺，其特征在于：在装置开工时，需要从不凝气补充管线向汽提塔内引入不凝气。
機械和訳	25.請求項 17 に記載のプロセスであって,その特徴は以下のとおりである:装置の起動時に,凝縮ガス補充管路から放散塔内に非凝縮ガスを導入する必要がある。
校正結果	25.請求項 17 に記載のプロセスであって,その特徴は以下のとおりである:装置の起動時に,非凝縮ガス補充管路からストリッピング塔内に非凝縮ガスを導入する必要がある。
解　説	原文の「不凝气补充管线」について、正しくは「非凝縮ガス補充管路」であるところ、校正前の機械和訳では「凝縮ガス補充管路」という逆の意味に訳されてしまっており、ミスリーディングである。原文の「不」の訳抜けと見なし、「重大な訳抜け」と判定すべきである。

原文中に存在する語句が翻訳文から欠落する「訳抜け」は、その性質上、重大な不備となることが多い。だが、その発生箇所によっては、内容理解にさほど重大な悪影響を与えない訳抜けも存在する。以下、一例を示す。

［重大と判定すべきでない訳抜け：CN107549443B］

原　文	2.如权利要求 1 所述的复合蛋白质起泡剂的制备方法，其特征在于，所述混合蛋白质溶液包括米糠蛋白溶液和卵白蛋白溶液，所述米糠蛋白溶液和卵白蛋白溶液的体积比为 1:1～3:1。
機械和訳	2.請求項 1 に記載の複合たんぱく質発泡剤の製造であって,その特徴は以下のとおりである:前記混合タンパク質溶液は米糠タンパク質溶液及び卵白アルブミン溶液を含み,前記米糠タンパク質溶液と卵白アルブミン溶液の体積比は 1:1～3:1 である。
校正結果	2.請求項 1 に記載の複合たんぱく質発泡剤の製造方法であって,その特徴は以下のとおりである:前記混合タンパク質溶液は米糠タンパク質溶液及び卵白アルブミン溶液を含み,前記米糠タンパク質溶液と卵白アルブミン溶液の体積比は 1:1～3:1 である。
解　説	本例は発泡剤を製造する方法の請求項であり、原文の「制备方法」は

	正しく「製造方法」と訳出する必要があるが、校正前の機械和訳はこれを「製造」とのみ訳しており、「方法」が訳抜けしている。ただし、発明を理解する上では致命的な誤解を生むものではなく、重大と見なすほどではない。

上例も、「重大とみなすべきでない訳抜け」の事例としては、やや「重大」寄りではある。だが、解説欄で述べたように、「方法」が抜けていても発明の実質的な内容理解にはほとんど悪影響はないため、「重大ではない」と判定するほうが妥当である。当然、本例よりも軽度な訳抜け、典型的な例としては「前記」や「該」が抜けているだけのものは、重大とは見なされない。

なお、「訳抜け」とは、本項冒頭で述べたとおり「原文中の要素が翻訳されず、翻訳文中から欠落する」ことを指す。適切な訳語に訳されていなかったとしても、何らかの訳語が出力されている限り、それは「訳抜け」ではなく、「誤訳」として扱う。

3．重大な湧き出しと、そうでないものの実例

「湧き出し」とは、原文中に存在しない要素が機械和訳中に出力されることを指す。こちらも「誤訳」や「訳抜け」と同じく、発明の内容理解に大きな悪影響を及ぼすもののみを重大と見なす。

［重大と見なすべき湧き出し：CN107710838B］

原　文	2.权利要求 1 所述的方法，其中所述 BS 通过正交投影最大化所述功率的利用。
機械和訳	2.請求項 1 に記載の方法であって,ここで前記 BS は直交投影最大化によって前記電力の利用を最大化する。
校正結果	2.請求項 1 に記載の方法であって,ここで前記 BS は直交投影によって前記電力の利用を最大化する。
解　説	原文の「最大化」は翻訳文末尾の「最大化する」に対応するが、機械和訳ではこれと重複して「直交投影最大化」と二重に訳出されてしまっている。こちらは原文中に対応する語が存在しないため、現象としては「湧き出し」となる。発明内容の正しい理解も阻害しており、重大と判定すべきケースである。

湧き出しの中にも、読み手が類推可能であり、発明の内容理解上さほど重大でないものも存在する。以下、「重大」と判定すべきでない湧き出しの実例を示す。

［重大と判定すべきでない湧き出し：CN107530341B］

原　文	4.如权利要求 1 所述的用途，其中所述癌症选自：结肠直肠癌、子宫内膜癌、多发性骨髓瘤、胰腺癌、肾癌和成胶质细胞瘤。
機械和訳	4.請求項 1 に記載の用途であって,その特徴は以下のとおりである:前記癌は以下から選択される:結腸直腸癌,子宮体癌,多発性骨髄腫,膵臓癌,腎臓癌及び神経膠芽腫からなる群から選択される。
校正結果	4.請求項 1 に記載の使用であって,ここで前記癌は以下から選択される:結腸直腸癌,子宮内膜癌,多発性骨髄腫,膵臓癌,腎臓癌及び神経膠芽腫。
解　説	校正前の機械和訳には、原文に存在しない「その特徴は以下のとおりである:」という文言が湧き出している。物理的には大きな湧き出しであるが、内容理解には影響がないため、「重大」とは判定しない。

「湧き出し」の重大性の判定も、「誤訳」や「訳抜け」と同様、その湧き出しが発生したことで発明の正しい内容理解に大きな悪影響があるか否かが基準である。上例は、事象としては明白な湧き出しであり、校正自体は必須である。だが、湧き出した文言があってもなくても発明の実質的な内容理解への影響はほとんどなく、そのため「重大な湧き出し」とは判定すべきでない。

なお、機械和訳では、翻訳文の文末に「2」など原文にない数値が出力されていることがある。「湧き出し」と見なすべき現象であるが、内容理解の悪影響という観点から「重大」と見なすべきかの判断は微妙である。本事業では、個人差を避けるため、このタイプの湧き出しについては重大と見なさない方針とした。

また、機械和訳では、原文の内容が不必要に重複出力されることがある。これらも「湧き出し」の一種である。だが、重複出力された内容が正しければ、単に「二度言っている」だけであり、発明内容への誤解は生じないことも多い。このような場合も、重大とは判定していない。

以　上

報告書 対応項番					3.2 用語の誤訳								3.3 訳抜け						3.4 湧き出し			3.5 数値の処理									3.6 英数字記号の処理				3.7 請求項の記載様式に係る不備			
不備の類型					3.2.1 学習データ中に存在しない語の誤訳	3.2.2 学習データ中に適訳語が存在する語の誤訳		3.2.3 係り受けの誤訳		3.2.4 事例集			3.3.1 用語レベルの訳抜け		3.3.2 入力文字数超過による訳抜け	3.3.3 セミコロン区切り文での訳抜け		3.3.4 事例集	3.4.1 一般的な湧き出し		3.4.2 不要な特許用語の湧き出し	3.5.1 文末への数値出力	3.5.2 漢数字の処理に係る不備			3.5.3 漢数字を含む化合物名称における誤訳	3.5.4 漢数字「一」の処理	3.5.5 事例集			3.6.1 数式	3.6.2 化合物名	3.6.3 事例集		3.7.1 請求項番号が本文の一部として扱われる不備	3.7.2 発明の主題を文末に出力する傾向		3.7.3 発明構成要素の訳ゆれ
					学習データに存在しない新語／マイナー語の誤訳	用語が正しい単位で認識されないことによる誤訳	訳語候補が複数存在することによる誤訳	不適切な助詞による係り受け誤り	不適切な語順による係り受け誤り	エンタメ系の誤訳	中国語の残存（未知語）	当て字の誤訳	難訳語の訳抜け	本来訳出できる語の突発的な訳抜け	長文の末尾で翻訳文が途絶する	セミコロン区切り文での翻訳単位レベルの大規模な訳抜け	請求項の途中で原文にない句点が入る	用語の羅列での訳抜け	訳抜けを伴わない湧き出し	訳抜けの補填的な湧き出し	不要な特許用語の湧き出し	訳抜け箇所の数字の羅列	漢数字から算用数字が派生	派生した数字が文脈に組み込まれる	「両」から算用数字2が派生	漢数字を含む化学物質名の誤訳	訳出すべき「一」の欠落／訳出すべきでない「一」の訳出	請求項番号の複雑な羅列	「S311」等の誤訳	「一一対応」など「一一」の誤訳	数式の反映ミス	英数字を含む化合物名の反映ミス	中国語を含む数式の誤訳	長大な英数字文字列の反映ミス	請求項番号が本文の一部として扱われる不備	主題を文末へ配置／再出力	主題を文末へ再出力した結果の誤訳	発明構成要素の訳ゆれ
					事例3.2.1-1、3.2.1-3、3.2.1-4、3.2.4-2、3.2.4-4、3.2.4-5	事例3.2.2-1	事例3.2.2-2、3.2.4-3	事例3.2.3-1	事例3.2.3-2	事例3.2.4-1	事例3.2.4-4	事例3.2.4-5	事例3.3.1-1	事例3.3.1-2	事例3.3.2-1、3.3.2-2	事例3.3.3-1	事例3.3.3-1	事例3.3.4-1	–	事例3.4.1-1、3.4.1-2、3.4.3-1、3.4.3-2	事例3.4.2-1、3.4.2-2	事例3.5.1-1、3.5.1-2	事例3.5.2-1、3.5.2-2、3.5.5-5	事例3.5.2-3、3.5.5-2、3.5.5-3	–	事例3.5.3-1、3.5.5-2	事例3.5.4-1、3.5.5-3、3.5.5-6	事例3.5.5-1	事例3.5.5-4	事例3.5.5-7	（事例3.6.1-1、3.6.1-2）	（事例3.6.2-1、3.6.2-2）	事例3.6.3-1	（事例3.6.3-2）	事例3.7.1-1、3.7.1-2	事例3.7.2-1、3.7.2-2、3.7.5-1	事例3.7.2-1、3.7.2-2、3.7.5-1	事例3.7.3-1
通番	ｾｸｼｮﾝ	ｻﾝﾌﾟﾙ番号	文献番号	文字数	3.2①	3.2②	3.2③	3.2④	3.2⑤	3.2⑥	3.2⑦	3.2⑧	3.3①	3.3②	3.3③	3.3④	3.3⑤	3.3⑥	3.4①	3.4②	3.4③	3.5①	3.5②	3.5③	3.5④	3.5⑤	3.5⑥	3.5⑦	3.5⑧	3.5⑨	3.6①	3.6②	3.6③	3.6④	3.7①	3.7②	3.7③	3.7④
1	A	A1	CN109170721B	1,350	1											1				1		1																
2	A	A2	CN107095875B	1,278				1																														
3	A	A3	CN105854194B	1,433												1	1		1			1																1
4	A	A4	CN105011755B	1,288			1										1																					1
5	A	A5	CN106901678B	1,271				1									1																			1	1	
6	A	A6	CN108653718B	833					1				1						1																1			
7	A	A7	CN108814718B	1,209		1	1		1		1													1	1													1
8	A	A8	CN106264384B	1,281					1												1															1		
9	A	A9	CN107308021B	1,136	1			1								1																						
10	A	A10	CN106061378B	1,528												1				1																1		1
11	A	A11	CN105748209B	1,352		1										1				1																		
12	A	A12	CN108835587B	1,020	1	1		1			1																											
13	A	A13	CN106852673B	1,331					1							1																						
14	A	A14	CN107820422B	1,387																			1													1		1
15	A	A15	CN107708656B	1,393	1									1					1				1	1												1		
16	A	A16	CN108697169B	2,740			1							1		1	1		1	1		1																1
17	A	A17	CN109198156B	2,140												1																						
18	A	A18	CN106616655B	1,210	1	1					1							1				1													1			
19	A	A19	CN107438418B	1,328																					1													
20	A	A20	CN107018925B	1,264	1																																	
21	B	B1	CN107053842B	881																	1				1													
22	B	B2	CN107054269B	1,243					1										1			1																1
23	B	B3	CN107984336B	1,339					1					1	1		1					1			1													
24	B	B4	CN106080312B	1,481					1																													
25	B	B5	CN106467078B	1,543					1							1				1																		
26	B	B6	CN105798914B	1,450		1	1																															1
27	B	B7	CN106270770B	1,169	1	1		1						1																								
28	B	B8	CN107521704B	1,202				1								1				1					1													1
29	B	B9	CN109689190B	1,612					1					1			1			1		1																
30	B	B10	CN108136341B	1,277										1			1																					
31	B	B11	CN106347309B	1,383					1																													
32	B	B12	CN106960486B	1,334	1		1																															
33	B	B13	CN106458012B	1,321												1	1			1																1		1
34	B	B14	CN106608276B	1,365				1	1					1						1																		
35	B	B15	CN108116315B	1,656																																		1
36	B	B16	CN107457784B	1,316			1		1							1																						
37	B	B17	CN106003134B	1,229		1							1				1								1													1
38	B	B18	CN106103327B	1,278							1																											
39	B	B19	CN109623854B	1,286																				1														
40	B	B20	CN108602186B	1,296					1					1		1																						
41	C	C1	CN105793437B	725										1																						1		
42	C	C2	CN105722956B	1,355										1			1																					1
43	C	C3	CN108706851B	1,294	1	1		1						1		1	1			1		1			1													1
44	C	C4	CN106459919B	1,681		1			1								1																					1
45	C	C5	CN108473956B	1,324	1									1					1	1	1																	
46	C	C6	CN109439056B	600				1																														
47	C	C7	CN105713592B	1,118	1									1			1		1	1																1	1	
48	C	C8	CN106103362B	1,337												1	1			1																		1
49	C	C9	CN103864106B	1,328				1						1			1																					1
50	C	C10	CN106318031B	841	1											1																						
51	C	C11	CN107227246B	1,307	1				1																													
52	C	C12	CN104379532B	1,299					1					1																						1	1	
53	C	C13	CN107056927B	1,309	1	1	1		1		1						1									1										1		
54	C	C14	CN106459677B	1,269					1								1																			1		
55	C	C15	CN109678964B	1,369	1											1						1																
56	C	C16	CN106676525B	1,241												1						1	1															
57	C	C17	CN105820036B	1,343	1				1																1													
58	C	C18	CN107236475B	1,284												1			1	1																1		
59	C	C19	CN106795392B	1,337												1	1			1	1	1		1														
60	C	C20	CN107011778B	1,313								1																										
61	D	D1	CN110485087B	1,271				1	1						1		1					1			1											1		
62	D	D2	CN109811484B	1,697	1		1		1				1												1													1
63	D	D3	CN109778499B	1,177												1						1																1
64	D	D4	CN108699755B	1,267										1			1		1																	1		

報告書 対応項番					3.2 用語の誤訳								3.3 訳抜け						3.4 湧き出し			3.5 数値の処理									3.6 英数字記号の処理				3.7 請求項の記載様式に係る不備			
不備の類型					3.2.1 学習データ中に存在しない語の誤訳	3.2.2 学習データ中に適訳語が存在する語の誤訳		3.2.3 係り受けの誤訳		3.2.4 事例集			3.3.1 用語レベルの訳抜け		3.3.2 入力文字数超過による訳抜け	3.3.3 セミコロン区切り文での訳抜け		3.3.4 事例集	3.4.1 一般的な湧き出し		3.4.2 不要な特許用語の湧き出し	3.5.1 文末への数値出力	3.5.2 漢数字の処理に係る不備			3.5.3 漢数字を含む化合物名称における誤訳	3.5.4 漢数字「一」の処理	3.5.5 事例集			3.6.1 数式	3.6.2 化合物名	3.6.3 事例集		3.7.1 請求項番号が本文の一部として扱われる不備	3.7.2 発明の主題を文末に出力する傾向		3.7.3 発明構成要素の訳ゆれ
					学習データに存在しない新語／マイナー語の誤訳	用語が正しい単位で認識されないことによる誤訳	訳語候補が複数存在することによる誤訳	不適切な助詞による係り受け誤り	不適切な語順による係り受け誤り	エンタメ系の誤訳	中国語の残存（未知語）	当て字の誤訳	難訳語の訳抜け	本来訳出できる語の突発的な訳抜け	長文の末尾で翻訳文が途絶する	セミコロン区切り文での翻訳単位レベルの大規模な訳抜け	請求項の途中で原文にない句点が入る	用語の羅列での訳抜け	訳抜けを伴わない湧き出し	訳抜けの補填的な湧き出し	不要な特許用語の湧き出し	訳抜け箇所の数字の羅列	漢数字から算用数字が派生	派生した数字が文脈に組み込まれる	「両」から算用数字2が派生	漢数字を含む化学物質名の誤訳	訳出すべき「一」の欠落／訳出すべきでない「一」の訳出	請求項番号の複雑な羅列	「S311」等の誤訳	「一一対応」など「一一」の誤訳	数式の反映ミス	英数字を含む化合物名の反映ミス	中国語を含む数式の誤訳	長大な英数字文字列の反映ミス	請求項番号が本文の一部として扱われる不備	主題を文末へ配置／再出力	主題を文末へ再出力した結果の誤訳	発明構成要素の訳ゆれ
					事例3.2.1-1、3.2.1-3、3.2.1-4、3.2.4-2、3.2.4-4、3.2.4-5	事例3.2.2-1	事例3.2.2-2、3.2.4-3	事例3.2.3-1	事例3.2.3-2	事例3.2.4-1	事例3.2.4-4	事例3.2.4-5	事例3.3.1-1	事例3.3.1-2	事例3.3.2-1、3.3.2-2	事例3.3.3-1	事例3.3.3-1	事例3.3.4-1	–	事例3.4.1-1、3.4.1-2、3.4.3-1、3.4.3-2	事例3.4.2-1、3.4.2-2	事例3.5.1-1、3.5.1-2	事例3.5.2-1、3.5.2-2、3.5.5-5	事例3.5.2-3、3.5.5-2、3.5.5-3	–	事例3.5.3-1、3.5.5-2	事例3.5.4-1、3.5.5-3、3.5.5-6	事例3.5.5-1	事例3.5.5-4	事例3.5.5-7	（事例3.6.1-1、3.6.1-2）	（事例3.6.2-1、3.6.2-2）	事例3.6.3-1	（事例3.6.3-2）	事例3.7.1-1、3.7.1-2	事例3.7.2-1、3.7.2-2、3.7.5-1	事例3.7.2-1、3.7.2-2、3.7.5-1	事例3.7.3-1
通番	ｾｸｼｮﾝ	ｻﾝﾌﾟﾙ番号	文献番号	文字数	3.2①	3.2②	3.2③	3.2④	3.2⑤	3.2⑥	3.2⑦	3.2⑧	3.3①	3.3②	3.3③	3.3④	3.3⑤	3.3⑥	3.4①	3.4②	3.4③	3.5①	3.5②	3.5③	3.5④	3.5⑤	3.5⑥	3.5⑦	3.5⑧	3.5⑨	3.6①	3.6②	3.6③	3.6④	3.7①	3.7②	3.7③	3.7④
65	D	D5	CN110651081B	1,312	1				1							1																				1		
66	D	D6	CN111850936B	1,313										1								1																1
67	D	D7	CN110552233B	1,257	1				1					1		1						1																
68	D	D8	CN110055784B	1,149	1				1		1				1	1				1					1													1
69	D	D9	CN109487393B	2,435	1	1	1	1	1						1	1	1					1	1			1									1			1
70	D	D10	CN109267291B	1,309			1	1	1							1	1					1														1		1
71	D	D11	CN109056389B	1,239	1	1			1				1	1									1															
72	D	D12	CN108677351B	1,309										1		1	1			1		1	1	1														
73	D	D13	CN106567224B	1,047	1	1																																1
74	D	D14	CN108103771B	1,377		1								1																								
75	D	D15	CN109778430B	1,429									1							1						1												
76	D	D16	CN107002357B	1,300	1			1	1												1		1	1		1												
77	D	D17	CN112323194B	1,391					1					1																								
78	D	D18	CN106592168B	1,130										1																								
79	D	D19	CN109280993B	1,300																						1												
80	D	D20	CN106948170B	1,942	1														1																			1
81	E	E1	CN110735600B	1,275				1	1		1		1	1	1		1																					
82	E	E2	CN108071378B	1,339																						1												
83	E	E3	CN108026723B	1,286					1					1									1	1	1		1									1		1
84	E	E4	CN110939119B	1,341	1				1		1				1	1						1													1			
85	E	E5	CN108999389B	1,253	1				1							1				1					1													
86	E	E6	CN110629893B	1,940	1	1		1	1		1		1	1					1																			
87	E	E7	CN109854253B	1,125	1				1		1					1	1																		1			
88	E	E8	CN109056457B	1,695	1																				1													
89	E	E9	CN110387872B	1,048	1				1					1		1																						
90	E	E10	CN109930634B	1,345					1							1	1																					1
91	E	E11	CN110453690B	1,314			1	1	1		1																											
92	E	E12	CN109681229B	1,383												1				1																		
93	E	E13	CN109183975B	1,252	1			1	1							1	1			1																1		1
94	E	E14	CN108086468B	505		1										1				1							1											1
95	E	E15	CN110820924B	1,301																															1			1
96	E	E16	CN108252671B	1,485								1																										1
97	E	E17	CN107013093B	1,343												1																						
98	E	E18	CN108442330B	1,243				1								1				1																		
99	E	E19	CN107143156B	1,078					1																1													
100	E	E20	CN108301713B	1,278											1																				1	1		
101	F	F1	CN109690086B	1,459																																		1
102	F	F2	CN105829819B	1,310		1																																
103	F	F3	CN111118874B	1,295												1	1										1											
104	F	F4	CN109154207B	1,458	1									1		1						1													1			1
105	F	F5	CN110529260B	1,294			1												1																			
106	F	F6	CN107709807B	2,340					1								1																		1	1		1
107	F	F7	CN107191967B	1,531		1										1						1																1
108	F	F8	CN110500188B	1,300					1							1				1																		
109	F	F9	CN105190047B	1,334			1		1							1	1			1																1	1	1
110	F	F10	CN110131073B	1,297					1					1		1	1			1																		
111	F	F11	CN110925835B	1,028	1																																	1
112	F	F12	CN105102823B	2,139											1				1			1													1	1		1
113	F	F13	CN109723559B	1,325																			1															
114	F	F14	CN108361343B	1,350					1																1													1
115	F	F15	CN106471314B	1,399										1					1																			
116	F	F16	CN107636280B	664										1																						1		
117	F	F17	CN110398070B	1,375										1		1				1																		1
118	F	F18	CN101893148B	1,513				1																														
119	F	F19	CN107560437B	1,750										1	1	1				1																		1
120	F	F20	CN107525301B	1,159										1													1											
121	G	G1	CN105869633B	1,709	1	1			1		1																											1
122	G	G2	CN108061851B	871										1						1																		
123	G	G3	CN107045238B	1,240										1						1		1														1		1
124	G	G4	CN105320800B	1,850		1			1				1						1																			
125	G	G5	CN108089419B	1,346	1									1			1					1					1											
126	G	G6	CN106257574B	1,288		1			1										1																	1		
127	G	G7	CN107391399B	1,109	1		1									1				1																		
128	G	G8	CN107066236B	2,074																			1															1

報告書 対応項番					3.2 用語の誤訳								3.3 訳抜け						3.4 湧き出し			3.5 数値の処理									3.6 英数字記号の処理				3.7 請求項の記載様式に係る不備			
不備の類型					3.2.1 学習データ中に存在しない語の誤訳	3.2.2 学習データ中に適訳語が存在する語の誤訳		3.2.3 係り受けの誤訳		3.2.4 事例集			3.3.1 用語レベルの訳抜け		3.3.2 入力文字数超過による訳抜け	3.3.3 セミコロン区切り文での訳抜け		3.3.4 事例集	3.4.1 一般的な湧き出し		3.4.2 不要な特許用語の湧き出し	3.5.1 文末への数値出力	3.5.2 漢数字の処理に係る不備			3.5.3 漢数字を含む化合物名称における誤訳	3.5.4 漢数字「一」の処理	3.5.5 事例集			3.6.1 数式	3.6.2 化合物名	3.6.3 事例集		3.7.1 請求項番号が本文の一部として扱われる不備	3.7.2 発明の主題を文末に出力する傾向		3.7.3 発明構成要素の訳ゆれ
					学習データに存在しない新語／マイナー語の誤訳	用語が正しい単位で認識されないことによる誤訳	訳語候補が複数存在することによる誤訳	不適切な助詞による係り受け誤り	不適切な語順による係り受け誤り	エンタメ系の誤訳	中国語の残存（未知語）	当て字の誤訳	難訳語の訳抜け	本来訳出できる語の突発的な訳抜け	長文の末尾で翻訳文が途絶する	セミコロン区切り文での翻訳単位レベルの大規模な訳抜け	請求項の途中で原文にない句点が入る	用語の羅列での訳抜け	訳抜けを伴わない湧き出し	訳抜けの補填的な湧き出し	不要な特許用語の湧き出し	訳抜け箇所の数字の羅列	漢数字から算用数字が派生	派生した数字が文脈に組み込まれる	「両」から算用数字2が派生	漢数字を含む化学物質名の誤訳	訳出すべき「一」の欠落／訳出すべきでない「一」の訳出	請求項番号の複雑な羅列	「S311」等の誤訳	「一一対応」など「一一」の誤訳	数式の反映ミス	英数字を含む化合物名の反映ミス	中国語を含む数式の誤訳	長大な英数字文字列の反映ミス	請求項番号が本文の一部として扱われる不備	主題を文末へ配置／再出力	主題を文末へ再出力した結果の誤訳	発明構成要素の訳ゆれ
					事例3.2.1-1、3.2.1-3、3.2.1-4、3.2.4-2、3.2.4-4、3.2.4-5	事例3.2.2-1	事例3.2.2-2、3.2.4-3	事例3.2.3-1	事例3.2.3-2	事例3.2.4-1	事例3.2.4-4	事例3.2.4-5	事例3.3.1-1	事例3.3.1-2	事例3.3.2-1、3.3.2-2	事例3.3.3-1	事例3.3.3-1	事例3.3.4-1	–	事例3.4.1-1、3.4.1-2、3.4.3-1、3.4.3-2	事例3.4.2-1、3.4.2-2	事例3.5.1-1、3.5.1-2	事例3.5.2-1、3.5.2-2、3.5.5-5	事例3.5.2-3、3.5.5-2、3.5.5-3	–	事例3.5.3-1、3.5.5-2	事例3.5.4-1、3.5.5-3、3.5.5-6	事例3.5.5-1	事例3.5.5-4	事例3.5.5-7	（事例3.6.1-1、3.6.1-2）	（事例3.6.2-1、3.6.2-2）	事例3.6.3-1	（事例3.6.3-2）	事例3.7.1-1、3.7.1-2	事例3.7.2-1、3.7.2-2、3.7.5-1	事例3.7.2-1、3.7.2-2、3.7.5-1	事例3.7.3-1
通番	ｾｸｼｮﾝ	ｻﾝﾌﾟﾙ番号	文献番号	文字数	3.2①	3.2②	3.2③	3.2④	3.2⑤	3.2⑥	3.2⑦	3.2⑧	3.3①	3.3②	3.3③	3.3④	3.3⑤	3.3⑥	3.4①	3.4②	3.4③	3.5①	3.5②	3.5③	3.5④	3.5⑤	3.5⑥	3.5⑦	3.5⑧	3.5⑨	3.6①	3.6②	3.6③	3.6④	3.7①	3.7②	3.7③	3.7④
129	G	G9	CN106249206B	1,376													1																					1
130	G	G10	CN106997353B	1,342		1			1																										1	1		
131	G	G11	CN105469330B	1,261		1			1										1																			
132	G	G12	CN108628723B	1,298					1										1																			
133	G	G13	CN105047198B	1,121										1		1				1																		
134	G	G14	CN107424022B	1,322			1									1																						
135	G	G15	CN106134269B	1,264		1			1										1																			
136	G	G16	CN106406531B	1,402			1		1		1					1				1																		
137	G	G17	CN108475030B	1,297																																		1
138	G	G18	CN105528392B	1,291			1										1																					1
139	G	G19	CN108132852B	1,159										1		1				1																		
140	G	G20	CN106250021B	1,316					1											1							1											
141	H	H1	CN105872250B	1,288		1			1				1		1				1	1		1	1															1
142	H	H2	CN104241266B	1,010	1				1						1		1		1						1													1
143	H	H3	CN107277850B	1,352			1		1																													
144	H	H4	CN105577961B	1,208	1												1																			1	1	
145	H	H5	CN107172270B	1,298					1																													
146	H	H6	CN106961250B	1,264													1																					
147	H	H7	CN103812750B	1,214		1										1				1																		
148	H	H8	CN103974348B	401	1																						1											
149	H	H9	CN107787553B	1,298					1					1																								
150	H	H10	CN109585452B	1,223		1								1		1				1	1									1								1
151	H	H11	CN105634675B	1,313																																		1
152	H	H12	CN105323143B	2,347	1		1	1	1					1							1														1	1	1	
153	H	H13	CN106059886B	1,353	1				1					1																						1	1	
154	H	H14	CN105379008B	1,478																																		
155	H	H15	CN108260363B	1,251										1			1		1				1			1												1
156	H	H16	CN106713110B	1,298										1		1				1																		
157	H	H17	CN107196692B	1,312	1									1			1				1	1													1			1
158	H	H18	CN106454683B	1,173	1																																	
159	H	H19	CN106658584B	1,284					1					1			1																					
160	H	H20	CN105655507B	1,266					1																					1								
合　計					44	27	19	21	63	0	13	2	9	47	11	53	40	1	22	39	8	26	12	7	17	7	7	0	0	2	0	0	0	0	13	29	7	53

他種エンジンによる機械和訳比較結果

報告書　掲載項番		事例番号	採取文献	誤訳の内容	原文	校正文	機械翻訳プラットフォーム	エンジン①	エンジン②	エンジン③	エンジン④	短評
3.2. 用語の誤訳	3.2.1. 学習データ中に存在しない語の誤訳	3.2.1-1	CN107124391B ｾｸｼｮﾝH	「硼单行为」誤訳	根据所述验证信息识别硼单行为，包括：识别所述无线接入点名称是否为随机字符串，若所述无线接入点名称为随机字符串，则确定所述客户端执行硼单行为；	前記認証情報に基づいて架空注文行動を識別することは,以下を含む:前記無線アクセスポイント名がランダム文字列であるかどうかを識別し,前記無線アクセスポイント名がランダム文字列と呼ばれる場合,前記クライアントが架空注文行動を実行したと判断する;	前記検証情報に基づいてブラシ単体行動を識別し,以下を含む:前記無線アクセスポイント名がランダム文字列であるかどうかを識別し,前記無線アクセスポイント名がランダム文字列と呼ばれる場合,前記クライアントがブラシシング行動を実行するかどうかを決定する;	前記検証情報に基づいてメニュー表示行為を識別することは,以下を含む:前記無線アクセスポイント名がランダム文字列であるか否かを識別し,前記無線アクセスポイント名がランダム文字列であれば,前記クライアント端末がチケット動作を実行することを決定する;	前記検証情報に基づいてブラシの挙動を識別することは,以下を含む:前記無線アクセスポイントの名称がランダム文字列であるか否かを識別し,前記無線アクセスポイントの名称がランダムな文字列である場合,前記クライアントがワンショットの挙動を実行することを決定する;	前記検証情報に基づいて、前記無線アクセスポイント名がランダム文字列であるか否かを識別するステップと、前記無線アクセスポイント名が前記ランダム文字列である場合、前記クライアントが前記スクラブ行を実行することを決定するステップと	検証情報による紙幣のスワイプ動作の識別には、無線アクセス ポイントの名前がランダムな文字列であるかどうかを識別し、無線アクセス ポイントの名前がランダムな文字列である場合は、クライアントが実行したと判断することが含まれます。紙幣をスワイプする動作;	全エンジン、訳語は異なるが何れも本来の意味には程遠い誤訳。エンジン③では訳抜けも発生している。
		3.2.1-2	CN112914750B ｾｸｼｮﾝA	「辅助口部防护装置」妥当訳 ※不備無し	2.根据权利要求1所述的一种瘫痪患者用辅助口部防护装置，其特征在于：所述下压板(1)与上压板(2)之间为活动连接，下压板(1)的下表面亦设计有牙套(3)。	2.請求項1に記載のてんかん患者用の補助的な口部保護装置であって,その特徴は以下のとおりである:前記下部押圧板(1)と上部押圧板(2)との間は可動接続され,下部押圧板(1)の下面にもマウスピース(3)が設計される。	2.請求項1に記載のてんかん患者用補助口部保護装置であって,その特徴は以下のとおりである:前記下押さえ板(1)と上押さえ板(2)との間が可動接続され,下押さえ板(1)の下面にもマウスピース(3)が設計される。	2. 請求項1に記載のてんかん患者用の補助口部保護装置であって、その特徴は以下のとおりである：前記下押さえ板(1)と上押さえ板(2)との間は可動的に接続され,下押さえ板(1)の下面にもマウスピース(3)が設計される。	2.請求項1に記載のてんかん患者用の補助口部保護装置であって,その特徴は以下のとおりである:前記下押さえ板(1)と上押さえ板(2)との間は可動接続され,下押さえ板(1)の下面にもマウスピース(3)が設計される。	2.請求項1に記載のてんかん患者用の補助口部防護装置であって,その特徴は以下のとおりである:前記下プラテン(1)と上プラテン(2)との間は可動接続され,下プラテン(1)の下面にもマウスピース(3)が設計される。	2. 下圧板(1)と上圧板(2)が柔軟に接続され、下圧板(1)の表面もブレース（3）で設計されています。	本例は機械翻訳プラットフォームで適切に訳されていた事例であるが、他種エンジンも、エンジン④を除けば、同様に適切に翻訳できている。一方、エンジン④の機械和訳は「輔助口部防护装置」を含む原文前半部分が完全に訳抜けしている。
		3.2.1-3	CN110927791B ｾｸｼｮﾝG	「深度学习」誤訳	7.根据权利要求1至6中任一项所述的方法，其特征在于，所述第一套深度学习网络和第二套深度学习网络基于卷积神经网络框架Caffe来构建，并采用Python编程语言来进行网络控制。	7.請求項1から6のいずれか一項に記載の方法であって,その特徴は以下のとおりである:前記第一セットのディープラーニングネットワークおよび第二セットのディープラーニングネットワークは,畳み込みニューラルネットワークフレームワークCaffeに基づいて構築され,且つPythonプログラミング言語を用いてネットワーク制御を行う。	7.請求項1から6のいずれか一項に記載の方法であって,その特徴は以下のとおりである:前記第一セットの深さ学習ネットワークおよび第二セットの深さ学習ネットワークは畳み込みニューラルネットワークフレームワークCaffeに基づいて構築され,且つPythonプログラミング言語を用いてネットワーク制御を行う。	7. 請求項1～6のいずれか1項に記載の方法であって、前記第一セットのディープラーニングネットワークおよび第二セットのディープラーニングネットワークは畳み込みニューラルネットワークフレームワークCaffeに基づいて構築され,且つPythonプログラミング言語を用いてネットワーク制御を行うことを特徴とする。	7.請求項1から6のいずれか一項に記載の方法であって,その特徴は,前記第一セットの深層学習ネットワークおよび第二セットの深層学習ネットワークが畳み込みニューラルネットワークフレームワークCaffeに基づいて構築され,且つPythonプログラミング言語を用いてネットワーク制御を行うことである。	7.請求項1～6のいずれか一項に記載の方法であって,その特徴は,前記第一深層学習ネットワークおよび第二深層学習ネットワークが畳み込みニューラルネットワークフレームワークCaffeに基づいて構築され,且つPythonプログラミング言語を用いてネットワーク制御を行うことである。	深層学習ネットワークの前記第１のセットおよび深層学習ネットワークの前記第２のセットは、畳み込みニューラルネットワークフレームワークＣａｆｆｅに基づいて構築され、ネットワーク制御のためにＰｙｔｈｏｎ言語でプログラムされる、請求項１から６のいずれか一項に記載の方法。	全エンジンとも「ディープラーニング」又は「深層学習」と適切に翻訳できている。
		3.2.1-4	CN108265104B ｾｸｼｮﾝC	「松锤研磨」「紧锤研磨」誤訳	4.根据权利要求1所述的构建方法，其特征在于，采用杜恩斯匀浆器对所述裂解产物进行松锤研磨不少于十次，紧锤研磨不少于十次，得到所述研磨产物。	4.請求項1に記載の構築方法であって,その特徴は以下のとおりである:ダウンス型ホモジナイザーを用いて前記溶解産物に対して粗調整用ベストルでの粉砕を10回以上行い,微調整用ベストルでの粉砕を10回以上行い,前記粉砕産物を得る。	4.請求項1に記載の構築方法であって,その特徴は以下のとおりである:ダウンスポホモジナイザーを用いて前記溶解生成物に対して10回以上のハンマーミリングを行い,ハンマーミリングが10回以上行われ,前記研磨製品が得られる。	4. 請求項1に記載の構築方法であって、ダンスホモジナイザーを用いて前記分解生成物に対して10回以上の松重り研磨を行い,10回以上の緊密重り研磨を行い,前記研磨生成物を得ることを特徴とする。	4.請求項1に記載の構築方法であって,その特徴は,ジンスホモジナイザーを用いて前記分解生成物に対してのンマー研磨を十回以上行い,ハンマーミルを十回以上粉砕し,前記研磨生成物を得ることである。	4.請求項1に記載の構築方法であって,その特徴は以下のとおりである:ダウンス型ホモジナイザーを用いて前記分解生成物に対して10回以上のルーズハンマー研磨を行い,10回以上のタイトハンマー研磨を行い,前記研磨生成物を得る。	ドーンズホモジナイザーを採用し、分解物を10回以上スラックハンマー粉砕し、10回以上タイトハンマー粉砕して粉砕物を得ることを特徴とする請求項1に記載の工法。.	機械翻訳プラットフォームで訳された「松锤研磨（粗調整用ベストルでの粉砕）」、「紧锤研磨（微調整用ベストルでの粉砕）」とも他の各エンジンでも誤訳された。特に「松锤／紧锤」の部分が難訳語であり、エンジン①では「松重り／緊密重り」、エンジン②は「ハンマー／ハンマー」と意味不明。これに比べればエンジン③の「ルーズハンマー／タイトハンマー」とエンジン④の「スラックハンマー／タイトハンマー」は、粗密の違いが表現できている分、いくらか良い。なお、エンジン④以外は「研磨（中国語）」の部分も「粉砕」ではなく「研磨」と誤訳している。
	3.2.2. 学習データ中に通訳語が存在する語の誤訳	3.2.2-1	CN113235177B ｾｸｼｮﾝD	「风湿度」誤訳	S3、通过侧吹风对复合纤维进行冷却成型，风温为13-15℃，风速为1-1.4m/s，风湿度为80-90%,冷却吹风位置在喷丝板下方40-60cm处；	S3,サイドブローにより複合繊維を冷却成形し,風温は13-15℃であり,風速は1-1.4m/sであり,風の湿度は80-90%であり,冷却送風位置は紡糸口金から40-60cm下の箇所である;	S3,サイドブローにより複合繊維を冷却成形し,風温は13-15℃であり,風速は1-1.4m/sであり,リウマチ度は80-90%であり,冷却送風位置は紡糸口金の下方40-60cmである;	S3、サイドブローによって複合繊維を冷却成形し,風温は13-15°Cであり,風速は1-1.4m/sであり,風湿度は80-90%であり,冷却ブロー位置は紡糸口金の下方40-60cmである;	S3サイドブローにより複合繊維を冷却して成形し,風温は13-15°Cであり,風速は1-1.4m/sであり,風湿度は80-90%であり,冷却ブロー位置は紡糸口金の下40-60cmに位置する;	S3、サイドブローにより複合繊維を冷却して成形し,風温は13-15℃であり,風速は1-1.4m/sであり,リウマチ度は80-90%であり,冷却ブロー位置は紡糸口金の下方40-60cmである;	S3. 横吹きで複合繊維を冷却成形し、風温13～15℃、風速1～1.4m/s、風湿度80～90%、冷却吹き付け位置40°-紡糸口金の下60cm;	エンジン③のみ、機械翻訳プラットフォームと同じ誤訳（「リウマチ度」）をしているが、その他のエンジンは正しく「風（の）密度」と訳せている。
		3.2.2-2	CN111945447B ｾｸｼｮﾝD	「上网」誤訳	4.根据权利要求1所述的一种基于聚丙烯酸型高吸水复合树脂的吸水纸的制备方法，其特征是：所述上网抄纸步骤中上网的浓度为0.3-1wt%，浆料的打浆度为28-42°，制得的湿面纤维层和湿底纤维层的含水率控制为55-70%。	4.請求項1に記載のポリアクリル酸系高吸水性複合樹脂に基づく吸水紙の製造方法であって,その特徴は以下のとおりである:前記ワイヤ上抄紙ステップにおけるワイヤ上の濃度は0.3-1wt%であり,スラリーの叩解度は28-42°であり,得られる湿潤表面繊維層と湿潤底面繊維層の含水率は55-70%に制御される。	4.請求項1に記載のポリアクリル酸系高吸水性複合樹脂に基づく吸水紙の製造方法であって,その特徴は以下のとおりである:前記ネットサーフィンステップにおけるインターネットの濃度は0.3-1wt%であり,パルプの叩解度は28-42°であり,製造された湿潤面繊維層と湿潤底繊維層の含水率は55-70%に制御される。	4. 請求項1に記載のポリアクリル酸型高吸水性複合樹脂に基づく吸水紙の製造方法であって、その特徴は以下のとおりである：前記上糸抄紙ステップにおける上糸の濃度は0.3-1wt%であり,パルプの叩解度は28-42°であり,製造された湿面繊維層及び湿底繊維層の含水率を55-70%に制御する。	4.請求項1に記載のポリアクリル酸型高吸水性複合樹脂に基づく吸水紙の製造方法であって,その特徴は以下のとおりである:前記上網抄紙工程における上網の濃度は0.3-1wt%であり,スラリーの叩解度は28-42°であり,製造された湿面繊維層及び湿底繊維層の含水率は55-70%に制御される。	4.請求項1に記載のポリアクリル酸型高吸水性複合樹脂に基づく吸水紙の製造方法であって,その特徴は以下のとおりである:前記トップワイヤ抄紙ステップにおけるトップワイヤの濃度は0.3-1wt%であり,パルプの叩解度は28-42°であり,製造されたウェット繊維層及びウェットフロア繊維層の含水率は55-70%に制御される。	【請求項４】 インターネット上で抄紙する工程におけるオンライン化濃度が０．３～１ｗｔｔであることを特徴とする請求項１～３のいずれかに記載のポリアクリル酸系吸水性複合樹脂系吸水紙の製造方法。%、スラリーの叩解温度は２８～４２°であり、調製された湿った表面繊維層および湿った底部繊維層の含水率は５５～７０％に制御される。	製紙分野案件で「上网」をインターネットと誤訳するか否かが焦点の事例。エンジン③の「トップワイヤ」は適訳、エンジン②の「上網」も意味は伝わるので正解とした。他のエンジンは「オンライン」「上糸」「インターネット」などいずれも誤訳。
	3.2.3. 係り受けの誤訳	3.2.3-1	CN105905272B ｾｸｼｮﾝB	係り受け誤訳	上平坦部(31)和下平坦部(35)，所述上平坦部(31)和所述下平坦部(35)形成与所述底板(17)的所述后边界(19)相邻接的弯折部(41)，所述上平坦部(31)与所述中心纵向轴线垂直，所述下平坦部(35)与所述底板(17)形成包括在30°与60°之间的角度；	上平坦部(31)と下平坦部(35)であって,前記上平坦部(31)と前記下平坦部(35)は前記底板(17)の前記後縁部(19)に隣接する屈曲部(41)を形成し,前記上平坦部(31)と前記中心縦軸は垂直であり,前記下平坦部(35)と前記底板(17)は30°と60°の間に含まれる角度を形成する;	上平坦部(31)と下平坦部(35)であって,前記上平坦部(31)と前記下平坦部(35)は前記底板(17)の前記後境界(19)に隣接する屈曲部(41)を形成し,前記上平坦部(31)と前記中央長手方向軸に垂直であり,前記下平坦部(35)と前記底板(17)は30°と60°の間に含まれる角度を形成する;	上平坦部(31)及び下平坦部(35)であって,前記上平坦部(31)と前記下平坦部(35)は前記底板(17)の前記後境界(19)に隣接する屈曲部(41)を形成し,前記上平坦部(31)は前記中心縦軸線に垂直であり,前記下平坦部(35)は前記底板(17)と30°と60°の間の角度を形成するもの;	上平坦部(31)と下平坦部(35),前記上平坦部(31)と前記下平坦部(35)は前記底板(17)の前記後境界(19)に隣接する屈曲部(41)を形成し,前記上平坦部(31)は前記中心縦方向軸線に垂直であり,前記下平坦部(35)と前記底板(17)は30°と60°の間に含まれる角度を形成する;	上部平坦部(31)及び下部平坦部(35)であって,前記上部平坦部(31)及び前記下部平坦部(35)は前記底板(17)の前記後境界(19)に隣接する屈曲部(41)を形成し,前記上部平坦部(31)は前記中心長手方向軸線に垂直であり,前記下部平坦部(35)は前記底板(17)と30°と60°との間に含まれる角度を形成するもの;	上部平坦部（３１）および下部平坦部（３５）、上部平坦部（３１）および下部平坦部（３５）は、底板（１７）の後部境界（１９）に隣接する接合部を形成し、屈曲部分（４１）、上部平坦部分（３１）は中心縦軸に垂直であり、下部平坦部分（３５）および底板（１７）は３０°から６０°の角度を含む角度を形成し、	機械翻訳プラットフォームでは不適切な助詞による係り受け誤りが生じた事例だが、他のエンジンは正しい係り受けで訳せている。
		3.2.3-2	CN107790355B ｾｸｼｮﾝB	係り受け誤訳	基板收容部，包含可相互连接分离的第1覆盖部及第2覆盖部，将支撑于所述支撑部的所述基板收容于所述第1覆盖部与所述第2覆盖部之间的处理空间；	基板収容部であって,互いに接続分離可能な第一被覆部及び第二被覆部を含み,前記支持部に支持された前記基板を前記第一被覆部と前記第二被覆部との間の処理空間に収容する;	基板収容部であって,互いに分離可能に分離された第1カバー部及び第2カバー部を含み,前記支持部に支持された前記基板を前記第1カバー部と前記第2カバー部との間に収容する処理空間;	互いに接続分離可能な第1被覆部および第2被覆部を含み、前記支持部に支持された前記基板を前記第1被覆部と前記第2被覆部との間の処理空間に収容する基板収容部;	基板収容部,互いに接続分離可能な第1被覆部および第2被覆部を含む,前記支持部に支持された前記基板を前記第1被覆部と前記第2被覆部との間の処理空間に収容する;	互いに接離可能な第1カバー部および第2カバー部を有し、前記支持部に支持された前記基板を前記第1カバー部と前記第2カバー部との間の処理空間に収容する基板収容部と	前記基板収納部は、連結及び分離可能な第１カバー部及び第２カバー部を含み、前記支持部に支持された前記基板を前記第１カバー部と前記第２カバー部空間との間に収納する工程と、	機械翻訳プラットフォームでは不適切な語順により係り受け誤りが生じた事例だが、他のエンジンはエンジン④以外は正しく訳せている。エンジン④は全体的に係り受けが乱れている。
3.2.4. 事例集		3.2.4-1	CN109943931B ｾｸｼｮﾝD	「打手」誤訳	清棉：抓棉打手转速为700～900r/min，开棉机梳针滚筒转速为300～500r/min，梳针滚筒与给棉罗拉之间的隔距10~13mm，综合打手转速为900~1000r/min，综合打手与天平之间隔距为8~12mm，棉卷罗拉转速为10~15r/min；	綿棉:ブラッキングピーターの回転数は700~900r/minであり,開綿機のキルシュナピーターの回転数は300~500r/minであり,キルシュナピーターとフィードローラとの間の距離は10~13mmであり,一体化ピーターの回転数は900~1000r/minであり,一体化ピーターと天秤との間隔距離は8~12mmであり,ラップローラの回転数は10~15r/minである;	綿:綿とんすけの回転数は700~900r/minであり,開綿機の針ドラムの回転数は300~500r/minであり,針ドラムとフィードローラとの間の距離は10~13mmであり,綿含とんすけ回転数は900~1000r/minであり,綿含とんすけと天秤との間隔距離は8~12mmであり,ロールローラの回転数は10~15r/minである;	清綿:綿掴みハンド回転数は700～900r/minであり,開綿機の櫛針ローラの回転数は300～500r/minであり,櫛針ローラと綿供給ローラとの間の間隔は10～13mmであり,綿含のハンド回転数は900～1000r/minであり,綿含ハンドと天秤との間の間隔は8～12mmであり,綿ロールローラの回転数は10～15r/minである;	綿清掃:綿掴みハンドの回転数は700～900r/minであり,綿解繊機の櫛針ローラの回転数は300～500r/minであり,櫛針ローラと綿供給ローラとの間の隔たり距離は10~13mmであり,綿含的な打手の回転数は900~1000r/minであり,綿含的な打撃と天秤との間の隔たり距離は8~12mmであり,綿ローラのローラの回転速度は10~15r/minである;	綿を除去する:綿掴みハンドルの回転速度は700～900r/minであり,開綿機の梳針ドラムの回転速度は300～500r/minであり,梳針ドラムと供給ローラとの間の間隔距離は10～13 mmであり,綿含ハンドルの回転速度は900～1000r/minであり,綿含ハンドルと天秤との間の間隔距離は8～12 mmであり,ラップローラの回転速度は10～15r/minである;	綿のクリーニング：綿ピッカーの回転速度は700～900r /分、綿切り機のカードシリンダーの回転速度は300～500r /分、カードシリンダーと綿送ローラーの間の距離は10～13mmです、および包括的なピーターの回転速度は900～1000r / min、ピーターとパランスの間の距離は8～12mm、ラップローラーの速度は10～15r / minです。	エンジン④の「ピーター」のみ正しい訳語が採用されているが、他は「ハンド」「ハンドル」と、エンタメ系の訳語ではないものの不適訳となっている。エンジン②の「打手」は、文字の意味からある程度伝わるが、中国語のまま出力されている可能性もあり、誤訳（未知語）とした（なお一箇所は「打撃」となっている）。
		3.2.4-2	CN109457541B ｾｸｼｮﾝD	「降焦烟用接装纸」誤訳、訳ゆれ	1.一种降焦烟用接装纸，其特征在于，包括内层和外层，所述内层和外层之间具有多个条状的第一支撑条和多个条状的第二支撑条，多个第一支撑条呈环形阵列分布，且相邻的第一支撑条间距相等，多个第二支撑条呈环形阵列分布，…	1.タール減少用チッピング用紙であって,その特徴は以下のとおりである:内層と外層を含み,前記内層と外層の間に複数のストリップ状の第一支持ストリップと複数のストリップ状の第二支持ストリップを有し,…	1.フォーカスタバコ用チッピング紙であって,その特徴は以下のとおりである:内層と外層を含み,前記内層と外層の間に複数のストリップ状の第一支持ストリップと複数のストリップ状の第2支持ストリップを有し,…	1. コークス低減タバコ用チップペーパーであって、内層及び外層を含み,前記内層と外層との間に複数のストライプ状の第一支持ストリップ及び複数のストライプ状の第二支持ストリップを有し,…	1,降コークスのタバコ用チッピングペーパー,その特徴は、内層および外層を含む,前記内層と外層との間に複数のストライプ状の第一支持バー及び複数のストライプ状の第二支持バーを有する,複数の第一支持ストリップは環状アレイ分布を呈する,…	1.タールスモーク低減用チッピング紙であって,その特徴は以下のとおりである:内層及び外層を含み,前記内層と外層との間に複数のストリップ状の第一支持ストリップ及び複数のストリップ状の第二支持ストリップを有し,…	1. 内層と外層からなり、内層の間に複数の帯状の第1支持帯と複数の帯状の第2支持帯があることを特徴とするタール低減シガレット用のチッピングペーパー。支持棒、複数の第１支持棒は円形配列で分配され、隣接する第１支持棒は等間隔で配置され、…	本件では「降焦烟用接装纸（タール減少用チッピング紙）」が発明の主題として全10項に記載されている。機械翻訳プラットフォームでは、うち2項で「タール低減用チッピング紙」と正しく訳されたが、残る8項は「タール用（の）チッピング用チッピング紙」（7項）、「フォーカスタバコ用チッピング紙」（1項）と、「降焦烟」の部分が誤訳されたり、「減少」の訳抜けと「チッピング」の埋め合わせ的湧き出しが発生したりしていた。 エンジン①は10項すべてで「コークス低減タバコ用チップペーパー」と訳されており、機械翻訳プラットフォームのような訳ゆれはないが「焦烟」を「コークス」と誤訳しており、品質は同レベル以下である。 ほぼ同じ学習データを使用しているエンジン②も「降焦烟」に対応できず、同じく「コークス低減」と誤訳したり、「タバコ降ろし」「降コークス」などの訳ゆれも発生しており、全項誤訳であった。 エンジン③は「タールスモーク低減用チッピング紙」など正しい意味が類推できる訳語が4項、「コークス煙低減用」など「降焦烟」部分が誤訳されているのが6項と、機械翻訳プラットフォームと同じような状況である。
					2.根据权利要求1所述的降焦烟用接装纸，其特征在于，所述第一、二、三、四海绵单元的弧度相等，在85-95度之间。	2.請求項1に記載のタール減少用チッピング紙であって,…	2.請求項1に記載のタール用のチッピング用チッピング紙であって,…	2. 請求項1に記載のコークス低減タバコ用チップペーパーであって、…	2.請求項1に記載のタバコ降ろし用のチッピング紙であって,…	2.請求項1に記載のタールタバコ用チッピング紙であって,…	…請求項１に記載のタールシガレット用のチッピングペーパー。	
					3.根据权利要求1所述的降焦烟用接装纸，其特征在于，多个第一支撑条的长度相等且相互平行，多个第二支撑条的长度相等且相互平行，且第一支撑条的长度等于第二支撑条的长度。	3.請求項1に記載のタール減少用チッピング紙であって,…	3.請求項1に記載のタール用チッピング用チッピング紙であって,…	3. 請求項1に記載のコークス低減タバコ用チップペーパーであって、…	3.請求項1に記載のタバコ降ろし用のチッピング紙であって,…	3.請求項1に記載のコークス煙低減用チッピング紙であって,…	…請求項１または２に記載のタール低減シガレット用チッピングペーパー。…	
					4.根据权利要求1所述的降焦烟用接装纸，其特征在于，多个第一支撑条的一端位于同一平面，多个第一支撑条的另一端位于同一平面，多个第二支撑条的一端位于同一平面，多个第二支撑条的另一端位于同一平面。	4.請求項1に記載のタール減少用チッピング紙であって,…	4.請求項1に記載のタール低減用チッピング紙であって,…	4. 請求項1に記載のコークス低減タバコ用チップペーパーであって、…	4.請求項1に記載のコークス低減タバコ用チッピング紙であって,…	4.請求項1に記載のコークス煙低減用のチッピング紙であって,…	前記複数の第1支持棒の一端が同一平面上に位置し、前記複数の第1支持棒の他端が同一平面上に位置し、前記複数の第1支持棒の他端が同一平面上に位置し、複数の第２支持棒は同一平面上に位置し、支持棒の一端は同一平面上に位置し、複数の第２支持棒の他端は同一平面上に位置する。 ※「降焦烟用接装纸」を含む原文前半部は訳出されず。	
					5.根据权利要求4所述的降焦烟用接装纸，其特征在于，每个第一支撑条和与其在同一直线的那个第二支撑条相距一段距离，所述一段距离等于第一透气单元沿着第一支撑条方向的长度。	5.請求項4に記載のタール減少用チッピング紙であって,…	5.請求項4に記載のタール用チッピング用チッピング紙であって,…	5. 請求項4に記載のコークス低減タバコ用チップペーパーであって、	5.請求項4に記載のコークス低減タバコ用チッピングペーパーであって,…	5.請求項4に記載のタール及び煙を低減するためのチッピング紙であって,…	…請求項４に記載のタール低減シガレット用のチッピングペーパー。…	
					6.根据权利要求1所述的降焦烟用接装纸，其特征在于，第一、二渗水孔的直径相等，均小于1mm。	6.請求項1に記載のタール減少用チッピング紙であって,…	6.請求項1に記載のタール用チッピング用チッピング紙であって,…	6. 請求項1に記載のコークス低減タバコ用チップペーパーであって、…	6.請求項1に記載のタバコ降ろし用のチッピング紙であって,…	6.請求項1に記載のコークス煙低減用チッピング紙であって,…	…請求項１～５のいずれか一項に記載のタール低減シガレット用チッピングペーパー。	

報告書	掲載項番	事例番号	採取文献	誤訳の内容	原文	校正文	機械翻訳プラットフォーム	エンジン①	エンジン②	エンジン③	エンジン④	短評
					7.根据权利要求1所述的降焦烟用接装纸，其特征在于，所述内层、外侧均为接装纸原纸。	7.請求項1に記載のタール減少用チッピング紙であって,…	7.請求項1に記載のタール用チッピング用チッピング紙であって,…	7. 請求項1に記載のコークス低減タバコ用チップペーパーであって、…	7.請求項1に記載のコークス低減タバコ用チッピング紙であって,…	7.請求項1に記載のタールタバコ低減用チッピング紙であって,…	…請求項１～６のいずれかに記載の減タールシガレット用チップペーパー。	エンジン④は訳ゆれが激しいものの「タール低減シガレット用のチッピングペーパー」など正しい意味が類推できる訳語が9項で採用されており（請求項4のみ訳抜け）、この用語に関しては最も優れた結果といえる。ただし、請求項4ではこの語を含む原文断章部が大規模に訳抜けしていたり、請求項6~8では引用する請求項の番号が湧き出しているなど別種の不備が見られた。
					8.根据权利要求1所述的降焦烟用接装纸，其特征在于，所述外层外侧具有水性油墨层。	8.請求項1に記載のタール減少用チッピング紙であって,…	8.請求項1に記載のタール用チッピング用チッピング紙であって,…	8. 請求項1に記載のコークス低減タバコ用チップペーパーであって、…	8.請求項1に記載のコークス低減タバコ用チッピング紙であって,…	8.請求項1に記載のタールタバコ低減用チッピング紙であって,…	…請求項１～７のいずれかに記載の減タールシガレット用チッピングペーパー。	
					9.根据权利要求1所述的降焦烟用接装纸，其特征在于，所述第一支撑条与内层和外层均粘结固定；所述第二支撑条与内层和外层均粘结固定。	9.請求項1に記載のタール減少用チッピング紙であって,…	9.請求項1に記載のタール用チッピング用チッピング紙であって,…	9. 請求項1に記載のコークス低減タバコ用チップペーパーであって、…	9.請求項1に記載のコークス低減タバコ用チッピング紙であって,…	9.請求項1に記載のコークス煙低減用チッピング紙であって,…	…請求項１に記載のタール低減シガレット用チッピングペーパー。…	
					10.根据权利要求1所述的降焦烟用接装纸，其特征在于，8个第二支撑条中有5个与第二海绵单元抵接，这5个第二支撑条中有2个第二支撑条与第三海绵单元也抵接，这5个第二支撑条中有3个第二支撑条与第三海绵单元不抵接。	10.請求項1に記載のタール減少用チッピング紙であって,…	10.請求項1に記載のタール低減用チッピング紙であって,…	10. 請求項1に記載のコークス低減タバコ用チップペーパーであって、…	10.請求項1に記載のタバコ降ろし用のチッピング紙であって,…	10.請求項1に記載の焦げ付き煙低減用チッピング紙であって,…	…請求項１に記載の減タールシガレット用チッピングペーパー。…	
		3.2.4-3	CN106579024B ｾｸｼｮﾝA	「研磨」誤訳	7.如权利要求1所述的梯度微粉碎同步热制骨粉和骨棚的加工方法，其特征在于，所述步骤四中还包括以下步骤：将初级骨粉制备成成品骨粉：经过离心喷雾后的初级骨粉再经过行星球磨机研磨至粒径为10～12μm的超细微的成品骨粉，其中，行星球磨机的转速为：450～500r/min，研磨时间为3～4h；之后，内层用洁净塑料袋，每袋按1Kg/袋进行包装得到骨粉产品，密闭入库，在干燥处阴凉处保存。	7.請求項1に記載の勾配微粉砕同期熟成骨粉及び骨ペーストの加工方法であって,その特徴は,前記ステップ四がさらに以下のステップを含むことである:一次骨粉を完成品骨粉に調製する:遠心噴霧した一次骨粉をさらに遊星ボールミル粉砕して粒径が10~12μmの超微細な完成品骨粉にし,ここで,遊星ボールミルの回転速度は以下のとおりである:450~500r/minであり,粉砕時間は3~4hである;その後,内層に清潔なプラスチック袋を用い,一袋ごとに1Kg/袋で包装して骨粉製品を得て,密封して入庫し,乾燥した日陰で保存する。	7.請求項1に記載の勾配微粉砕同期熟成骨粉と骨ペーストの加工方法であって,その特徴は以下のとおりである:前記ステップ4はさらに以下のステップを含む:一次骨粉を完成品骨粉に調製する:遠心噴霧した一次骨粉をさらに遊星ボールミル粉砕して粒径が10から12μmの超微細な完成品骨粉にし,ここで,遊星ボールミルの回転速度は以下のとおりである:450から500r/minであり,研磨時間は3から4hである;その後,内層は清潔なプラスチック袋であり,1袋あたり1Kg/袋で包装して骨粉製品を得て,密閉して保管し,乾燥箇所で日陰で保存する。	7. 請求項1に記載の勾配微粉砕同期熟成骨粉及び骨ペーストの加工方法であって、前記ステップ4はさらに以下のステップを含むことを特徴とする：一次骨粉を完成品の骨粉に調製する:遠心噴霧後の一次骨粉を遊星ボールミルで粒径が10～12μmの超微細な完成品骨粉に粉砕し,ここで,遊星ボールミルの回転速度は以下のとおりである:450～500r/min,研磨時間は3～4hである;その後,内層は清潔なプラスチック袋を用い,各袋を1Kg/袋で包装して骨粉製品を得て,密閉して入庫し,乾燥箇所の冷暗所で保存する。	7.請求項1に記載の勾配微粉砕を同期して熟成した骨粉と骨ペーストの加工方法であって、前記ステップ4がさらに以下のステップを含むことである:一次骨粉を完成品の骨粉に調製する:遠心噴霧を経た後の一次骨粉を遊星ボールミルで粒径が10~12μmの超微細な完成骨粉に粉砕し,ここで,遊星ボールミルの回転速度は:450~500r/minであり,粉砕時間は3~4hである;その後,内層はクリーンなプラスチック袋を用い,各袋は1Kg/袋で包装して骨粉製品を得て,密閉して入庫し,乾燥部の涼しい場所に保存する。	7.請求項1に記載の勾配微粉砕による同時調理骨粉及び骨ペーストの加工方法であって,その特徴は,前記ステップ4がさらに以下のステップを含むことである:一次骨粉を製品骨粉に製造する:遠心噴霧した後の一次骨粉をさらに遊星ボールミルで粒径が10～12μmの超微細な製品骨粉に粉砕し,ここで,遊星ボールミルの回転速度は:450～500r/minであり,粉砕時間は3～4hである;その後,内層は清潔なビニール袋を用い,各袋を1Kg/袋で包装して骨粉製品を得て,密閉して入庫し,乾燥した場所の冷暗所で保存する。	【請求項７】 前記ステップ４は、次のステップも含むことを特徴とする、請求項１に記載の同時加熱された骨粉末および骨ペーストの勾配微粉砕処理方法。骨粉を遊星ボールミルで粉砕し、粒子径 10 ～ 12 μｍの超微粒骨粉に仕上げます. 遊星ボールミルの速度は 450 ～ 500 r/min、粉砕時間は３～４時間後、内層をきれいなビニール袋で覆い、１袋あたり 1Kg で包装して骨粉製品とし、密閉して保管し、乾燥した涼しい場所に保管します。	中国語「研磨」を適切に訳し分け（研磨／粉砕）できるかが焦点の事例。文中に二度出現し、いずれも「粉砕」と訳すのが正解だが、エンジン①のみ機械翻訳プラットフォームと同様、2回目を「研磨」と誤訳してしまっている。他のエンジンは二回とも「粉砕」と適切に翻訳できている。
		3.2.4-4	CN107019420B ｾｸｼｮﾝA	「螺纹举升机构」一部未知語	所述限压阀(50)和所述锅盖(20)之间设置有螺纹举升机构(70)，所述电机(81)驱动所述限压阀(50)旋转时，所述限压阀(50)在所述螺纹举升机构(70)的作用下上升或者下降，从面开启或者封闭所述排气通道(30)，	前記圧力制限弁(50)と前記鍋蓋(20)との間にねじリフト機構(70)が設けられ,前記モータ(81)が前記圧力制限弁(50)の回転を駆動する時,前記圧力制限弁(50)は前記ねじリフト機構(70)の作用で上昇又は下降し,それにより前記排気通路(30)を開閉し,	前記圧力制限弁(50)と前記鍋蓋(20)との間にねじ纹举昇降機構(70)が設置され,前記モータ(81)が前記圧力制限弁(50)を回転駆動すると,前記圧力制限弁(50)は前記ねじ纹举昇降機構(70)の作用で上昇又は下降し,それにより前記排気通路(30)を開閉し	前記圧力制限弁(50)と前記鍋蓋(20)との間にネジリフト機構(70)が設置され,前記モータ(81)が前記圧力制限弁(50)の回転を駆動する時,前記圧力制限弁(50)は前記ネジリフト機構(70)の作用で上昇又は下降し,それにより前記排気通路(30)を開閉し、	前記圧力制限弁(50)と前記鍋蓋(20)との間にねじ挙上機構(70)が設けられ,前記モータ(81)は前記圧力制限弁(50)の回転を駆動する時,前記圧力制限弁(50)は前記ねじ挙上機構(70)の作用で上昇又は下降し,それにより前記排気通路(30)をオン又はオフにし,	前記圧力制限弁(50)と前記鍋蓋(20)との間にネジリフト機構(70)が設置され,前記モータ(81)が前記圧力制限弁(50)を駆動して回転させる時,前記圧力制限弁(50)は前記ネジリフト機構(70)の作用で上昇又は下降し,それにより前記排気通路(30)を開放又は密閉し	圧力制限弁(50)と鍋蓋(20)との間にはねじ込み式の昇降機構(70)が配置されており、モータ(81)が圧力制限弁(50)を回転駆動すると、圧力制限弁(50)が回転す る。）は、糸リフト機構（７０）の作用により上昇または下降し、それによって排気通路（３０）を開閉し、	「螺纹举升机构」という用語について、機械翻訳プラットフォームが「ねじ纹举昇降機構」と中国語のまま訳出してしまった事例。エンジン①は「ネジリフト機構」、エンジン③は「ネジリフト機構」とほぼ正解訳となっている。エンジン②の「ねじ挙上機構」も不自然な日本語だが指す意味は大差なく、正解とした。エンジン④は１回目の「ねじ込み式の昇降機構」は意味が類推できるため正解、ただし2回目は「糸リフト機構」と訳ゆれしており、かつ意味的にも不正解。
		3.2.4-5	CN112832018B ｾｸｼｮﾝD	当て字の誤訳	所述的神经性化学战剂为沙林、塔崩、梭曼、维埃克斯或环沙林；	前記神経性化学兵器はサリン,タブン,ソマン,ブイエクス又はシクロサリンである;	前記神経因性ケミカル兵器はサリン、タワー崩壊、ゼーマン、ベルル又はシクラムリンである;	前記神経性化学戦薬はサリン、タワー崩壊、クロマン、ビールクス又はシクロサリンである;	前記神経性化学兵器はサリン、タッビング、シャトルマン、ウィルエクス又はシクロサリンである;	前記神経化学兵器剤はサリン、タブン、ソマン、ベキサン又はシクロサリンである;	神経化学兵器は、サリン、タブン、ソマン、ビエックス、またはシクロサリンです。	エンジン①とエンジン②は機械翻訳プラットフォームとほぼ同等の誤訳が生じている。一方、エンジン③とエンジン④は、「維埃克斯：ブイエックス」がそれぞれ「ベキサン」「ビエックス」となっているが、異表記の範疇とも見なせ、ほぼ正解といえる。
3.3. 訳抜け	3.3.1. 用語レベルの訳抜け	3.3.1-1	CN105574733B ｾｸｼｮﾝG	「当期销售器」訳抜け	当期销售器，用于根据初始采购量商品采购条款发布当期线下线上销售条款；	現時販売器であって,初期購入量と商品購入条件に基づいて,現在のオフライン及びオンラインの販売条件を公開するために用いられる;	初期仕込み量商品調達条項に基づいてオンラインオフラインライン上販売条件を発行するために用いられる;	当期販売器であって,初期購入量商品の購入条項に基づいて当期のオフライン販売条項を発行するために用いられるもの;	相期販売機であって,初期採得量の商品の購入代金に基づいて現時点のオフライン販売代金を発行することに用いられるもの;	前記初期仕込み量製品購入条件に従って、現在のオフライン販売条件を発行するように構成された現在の販売者と	現在の販売デバイスを使用して、初回購入量の商品購入条件に従って、現在のオフラインおよびオンラインの販売条件を公開します。	マイナー用語で機械翻訳プラットフォームでは訳抜けした「当期销售器：現時販売器」だが、エンジン①とエンジン④はそれぞれ「当期販売器」「現在の販売装置」と例とか意味を類推可能な訳語で訳出されている。一方、エンジン②とエンジン④は訳出自体はされたがそれぞれ「現在の販売者」「相期販売機」とかなり意味がずれており、正解とは見なせない。
					预售销售器，用于发布下一个销售周期线上商品不限量预售条款。	先行販売用販売器であって,次の販売サイクルのオンライン商品における無制限の先行販売条件を公開するために用いられる。	販売売り手は,次の販売周期ライン上の商品が無制限に販売される条項を発行する。	販売予定販売器であって,次の販売周期線における商品の非制限量の販売予定条項を発行するために用いられるもの。	販売機は、次の販売サイクルライン上の商品の無限大販売代金を発行するために使用される。	販売促進装置であって,次の販売サイクルラインにおける商品の無制限の販売促進条項を発行することに用いられるもの。	先行販売者は、次の販売サイクルでオンライン製品の無制限の先行販売条件を公開するために使用します。	上行の続き。こちらもマイナー用語「預售销售器：先行販売用販売器」が存在し、機械翻訳プラットフォームでは「販売売り手」とニュアンスが大幅に不足した訳語となっている。各エンジンにおいても、「販売予定販売器」「販売機」「販売促進装置」「先行販売者」と、いくらかベターではあるが依然として誤訳された。
		3.3.1-2	CN106459922B ｾｸｼｮﾝC	「细胞表面标记」突発的訳抜け	1.一种心肌前体细胞，其是CD82为阳性，CD73和CD105为阴性，且选自ROR2、CD13、PDGFRα和KDR中的至少一个细胞表面标记为阳性的细胞。	1.心筋前駆細胞であって,それはCD82が陽性であり,CD73及びCD105が陰性であり,且つROR2,CD13,PDGFRα及びKDRからなる群から選択される少なくとも一つの細胞表面マーカーが陽性の細胞である。	1.心筋前駆細胞であって,それはCD82が陽性であり,CD73及びCD105が陰性であり,且つROR2,CD13,PDGFRα及びKDRからなる群から選択される少なくとも一つのマーカーが陽性である細胞である。	1. 心筋前駆細胞であって、これは、CD82が陽性であり、CD73およびCD105が陰性であり、ROR2、CD13、PDGFRαおよびKDRからなる群から選択される少なくとも1つの細胞表面マーカーが陽性である細胞である。	1.CD82が陽性であり、CD73およびCD105が陰性であり、ROR2、CD13、PDGFRαおよびKDRから選択される少なくとも1つの細胞表面が陽性である細胞である心筋前駆細胞。	CD82が陽性であり、CD73およびCD105が陰性であり、かつROR2、CD13、PDGFRαおよびKDRからなる群より選択される少なくとも1つの細胞表面マーカーが陽性である、心筋前駆細胞。	１．ＣＤ８２陽性、ＣＤ７３およびＣＤ１０５陰性、ならびにＲＯＲ２、ＣＤ１３、ＰＤＧＦＲαおよびＫＤＲから選択されるかなくとも１つの細胞表面マーカーが陽性である心筋前駆細胞。	機械翻訳プラットフォームでは、請求項1のみ、他の文で問題なく訳せていた「細胞表面標記，細胞表面マーカー」が「マーカー」と部分的に訳抜けした。エンジン②も同じ個所で「細胞表面」と逆方向の訳抜けが発生しており、請求項1でのこの語の使われ方が訳抜けを誘発しやすい何らかの条件を有している可能性がある。ただし、その他のエンジンでは請求項1、請求項3とも問題なく訳出されており、突発性の訳抜けである可能性も高い。 なお、エンジン③の請求項3の訳文は、原文末尾の「細胞マーカーB群」及び同「C群」のマーカー名称の羅列を本文に組み入れ、「細胞マーカーB群」と同「C群」の繰り返しを避けている。原文中には「細胞表面标记」が6回出現するのに対し、機械翻訳文には「細胞表面マーカー」が4回しか出現しないが、これは訳抜けではなく上記省略によるものであり、不備とは判定しなかった。 また、エンジン④では、「請求項１に記載の」であるべきところが「請求項１または２」となっており、別種の不備が生じている。
		3.3.1-2	CN106459922B ｾｸｼｮﾝC		3.根据权利要求1所述的心肌前体细胞，其是选自下面的细胞表面标记B组中的至少一个细胞表面标记为阴性、和/或选自下面的细胞表面标记C组中的至少一个细胞表面标记为阳性的细胞，细胞表面标记B组：CD7、CD37、CD43、CD144、STRO-1、CD177、CD163、细胞表面标记C组：CD137L、CD140b、CD180、CD252、CD344、CD118、CD99R。	3.請求項1に記載の心筋前駆細胞であって,それは以下の細胞表面マーカーB群から選択される少なくとも一つの細胞表面マーカーが陰性であり,及び/又は以下の細胞表面マーカーC群から選択される少なくとも一つの細胞表面マーカーが陽性の細胞であり,細胞表面マーカーB群:CD7,CD37,CD43,CD144,STRO－1,CD177,CD163,細胞表面マーカーC群:CD137L,CD140b,CD180,CD252,CD344,CD118,CD99R。	3.請求項1に記載の心筋前駆細胞であって,それは以下の細胞表面マーカーB群から選択される少なくとも一つの細胞表面マーカーが陰性であり,及び/又は以下の細胞表面マーカーC群から選択される少なくとも一つの細胞表面マーカーが陽性である細胞であり細胞表面マーカーB群:CD7,CD37,CD43,CD144,STRO－1,CD177,CD163細胞表面マーカーC群:CD137L,CD140b,CD180,CD252,CD344,CD118,CD99R。	3. 請求項1に記載の心筋前駆細胞であって、以下の細胞表面マーカーB群から選択される少なくとも一つの細胞表面マーカーが陰性であり、及び/又は以下の細胞表面マーカーC群から選択される少なくとも一つの細胞表面マーカーが陽性である細胞であり,細胞表面マーカーB群である：CD7、CD37、CD43、CD144、STRO-1、CD177、CD163,細胞表面マーカーC群: CD137L、CD140b、CD180、CD252、CD344、CD118 、CD99R。	3.以下の細胞表面マーカーB群から選択される少なくとも1つの細胞表面マーカーが陰性であり、及び/又は、以下の細胞表面マーカーC群から選択される少なくとも1つの細胞表面マーカーが陽性である細胞であり、細胞表面マーカーB群:CD7、CD37、CD43、CD144、STRO-1、CD177、CD163、細胞表面マーカーC群:CD137L、CD140b、CD180、CD252、CD344、CD118、CD99Rから選択される、請求項1に記載の心筋前駆細胞。	以下の細胞表面マーカーB群:CD7、CD37、CD43、CD144、STRO-1、CD177、CD163から選択される少なくとも1つの細胞表面マーカーが陰性であり、かつ/または以下の細胞表面マーカーC群:CD137L、CD140b、CD180、CD252、CD344、CD118、CD99Rから選択される少なくとも1つの細胞表面マーカーが陽性である細胞である、請求項1に記載の心筋前駆細胞。	下記細胞表面マーカー群Bから選択される少なくとも1つの細胞表面マーカーが陰性であり、および/または下記細胞表面マーカーC群A細胞から選択される少なくとも1つの細胞表面マーカーが陰性である、請求項1または2に記載の心筋前駆細胞。陽性の細胞表面マーカー、細胞表面マーカー グループ B: CD7、CD37、CD43、CD144、STRO-1、CD177、CD163、細胞表面マーカー グループ C: CD137L、CD140b、CD180、CD252、CD344、CD118、CD99R。	

他種エンジンによる機械和訳比較結果

報告書　掲載項番		事例番号	採取文献	誤訳の内容	原文	校正文	機械翻訳プラットフォーム	エンジン①	エンジン②	エンジン③	エンジン④	短評
	3.3.2. 入力文字数超過による訳抜け	3.3.2-1	CN107846845B ｾｸｼｮﾝC	長文末尾での翻訳文途絶①	7.根据权利要求1至4中任一项所述的捆包机，其中所述捆包机(1)设置有运动装置(99、85、87、185、187)，所述运动装置用于使为门的形式的所述进料器管道(23)的底壁(37)的第一门(59)和也为门的形式的所述进料器管道(23)的底壁(37)的第二门(61)选择性地在关闭位置和部分打开位置之间运动，在所述关闭位置中，所述第一门(59)的与其附接到所述输送通道的底壁(51)的边缘相对的边缘(67)与所述进料器管道(23)的底壁(37)的第二门(61)的边缘(69)相邻，由此与所述进料器管道(23)的底壁(37)的第二门(61)一起提供所述下边界的至少一部分，在所述部分打开位置中，所述进料器管道(23)的底壁(37)的第一门(59)和底壁(37)的第二门(61)远离所述进料器管道(23)的顶壁(35)定位，由此使得所述下边界保持完整。	7. 請求項1から4のいずれか一項に記載の梱包機であって,ここで前記梱包機(1)に運動装置(99,85,87,185,187)が設けられ,前記運動装置はドアの形態の前記フィーダーパイプ(23)の底壁(37)の第一ドア(59)及びドアの形態の前記フィーダーパイプ(23)の底壁(37)の第二ドア(61)を選択的に閉位置と一部の開位置の間で運動させるために用いられ,前記閉位置において,前記第一ドア(59)のそれが前記搬送通路の底壁(51)に取り付けられたエッジに対向するエッジ(67)は前記フィーダーパイプ(23)の底壁(37)の第二ドア(61)のエッジ(69)に隣接し,これにより前記フィーダーパイプ(23)の底壁(37)の第二ドア(61)と共に前記下部境界の少なくとも一部を提供し,前記一部の開位置において,前記フィーダーパイプ(23)の底壁(37)の第一ドア(59)及び底壁(37)の第二ドア(61)は前記フィーダーパイプ(23)の頂壁(35)から離れるように位置決めされ,これにより下側境界が完全に保持する。	7.請求項1～4のいずれか一項に記載の梱包装置は,梱包装置(1)運動装置(99,85,87,185,187)が設けられており,前記運動装置使为门形態用の前記供給管(23)の底壁(37)の第一の扉(59)とも,ドアの形の前記供給パイプ(23)の底壁(37)の第二のドア(61)の閉位置と開位置との間で選択的に移動可能であり,前記閉鎖位置において,前記第一のゲート(59)の前記供給通路の底壁(51)に取り付けられている端部とは反対側の縁部(67)は,供給導管(23)の底(37)の第二(61)の縁部(69)に隣接して配置され,これにより,前記供給パイプ(23)の底壁(37)の第二ゲート(61)と前記下側境界の少なくとも一部を提供する,部分的に開いた位置では,前記供給管(23)の前記底壁(37)の第一	7. 請求項1～4のいずれか1項に記載の梱包機であって、前記梱包機(1)に運動装置(99、85、87、185、187)が設けられ,前記運動装置はドアの形態の前記フィーダダクト(23)の底壁(37)の第一ドア(59)及びドアの形態の前記フィーダダクト(23)の底壁(37)の第二ドア(61)を選択的に閉位置と部分開位置との間で運動させるために用いられ,前記閉位置において,前記第一ドア(59)の前記搬送通路の底壁(51)に取り付けられた縁部に対向する縁部(67)は前記フィーダダクト(23)の底壁(37)の第二ドア(61)の縁部(69)に隣接し,それにより前記フィーダダクト(23)の底壁(37)の第二ドア(61)と共に前記下境界の少なくとも一部を提供し,前記部分開位置において,前記フィーダダクト(23)の底壁(37)の第一ドア(59)及び底壁(37)の第二ドア(61)は前記フィーダダクト(23)の上壁(35)から離れて位置し、これにより、下側境界が無傷のままとなる。	7.請求項1~4のいずれか一項に記載の梱包機,前記梱包機(1)に運動装置(99、85、87、185、187)が設置される,移動装置は、ドアの形態のフィーダダクト(23)の底壁(37)の第1のドア(59)と、ドアの形態のフィーダダクト(23)の底壁(37)の第2のドア(61)とを、閉鎖位置と部分開放位置との間で選択的に移動させるために使用される。,前記閉位置において、前記第1ドア(59)の前記搬送通路の底壁(51)に取り付けられた縁に対向する縁(67)は、前記フィーダダクト(23)の底壁(37)の第2ドア(61)の縁(69)に隣接している,これにより、前記フィーダーダクト(23)の底壁(37)の第2扉(61)とともに前記下境界の少なくとも一部が提供されている,前記部分開放位置において、フィーダーダクト(23)の底壁(37)の第1のドア(59)および底壁(37)の第2のドア(61)がフィーダーダクト(23)の頂壁(35)から離れて位置する,これにより、下側境界は無傷のままである。	前記ケーサー(1)は、ドアの形態の前記フィーダダクト(23)の前記底壁(37)の第1のドア(59)と、同じくドアの形態の前記フィーダダクト(23)の前記底壁(37)の第2のドア(61)とを、前記搬送チャネルの前記底壁(51)に取り付けられた縁部とは反対側の前記第1のドア(59)の縁部(67)が前記フィーダダクト(23)の前記底壁(37)の前記第2のドア(61)の縁部(69)に隣接し、それによって前記フィーダダクト(23)の前記底壁(37)の前記第2のドア(61)と共に前記下側境界の少なくとも一部を提供する閉鎖位置と、前記フィーダダクト(23)の前記底壁(37)の前記第1のドア(59)および前記底壁(37)の前記第2のドア(61)が前記フィーダダクト(23)の前記上壁(35)から離れて位置し、それによって前記下側境界を無傷のままにする部分的開放位置との間で選択的に移動させるための移動手段(99、85、87、185、187)を備える、請求項1から6のいずれか一項に記載のケーサー。	前記結束機（１）が、前記底壁の第１のゲート（５９）を作るための運動手段（９９、８５、８７、１８５、１８７）を備えている、請求項１から４のいずれか一項に記載の結束機。ゲートの形態の前記フィーダダクト（２３）の（３７）と、ゲートの形態の前記フィーダダクト（２３）の底壁（３７）の第１のゲート（５９）と、第２のドア（６１）と、閉鎖位置と、第１のドア（５９）が送達チャネルの底壁（５１）に取り付けられる部分的開放位置との間で選択的に移動可能である.前記フィーダーダクト（２３）の底壁（３７）の第２のドア（６１）、それによって前記フィーダーダクト（２３）に隣接するフィーダーダクト（２３）の底壁（３７）の第２のドア（６１））は、フィーダーダクト（２３）（５９）の底壁（３７）の第１のドアと底壁の第２のドア（６１）とを合わせて、部分的に開いた位置で、下部境界の少なくとも一部を提供する。（３７）は、供給ダクト（２３）の頂壁（３５）から離れて配置され、それによって下側境界をそのまま残す。	機械翻訳プラットフォームで翻訳文が途中で途絶した事例。他のエンジンでは翻訳文は文末まで出力されている。係り受けが正しく訳せているのはエンジン①と②だが、エンジン②は文が一節ごとに終止形で訳され、日本語文として不自然となっている。一方、エンジン③と④には係り受け誤り（赤字部分）が生じており、特にエンジン④は非常に不正確。
		3.3.2-2	CN107505041B ｾｸｼｮﾝG	長文末尾での翻訳文途絶②	1.基于相位敏感光时域反射计的相位解调方法，其特征在于，该方法基于相位解调装置，所述相位解调装置包括窄线宽激光器、第一光耦合器、光调制器、光放大器，环形器、第二光耦合器、传感光纤、脉冲信号发生器、光电探测器和数据采集器，所述传感光纤中均匀间隔刻有弱反射光栅，第二光耦合器为50:50光耦合器，脉冲信号发生器用于控制光调制器和数据采集器；窄线宽激光器发出连续光，该连续光经第一耦合器分成功率不等的两路光，其中功率较小的一路光作为本地参考光，而功率较大的一路光作为探测光，该探测光经光调制器调制成探测双脉冲，该探测双脉冲经光放大器放大后输入光环行器的第一端口，并由光环行器的第二端口注入传感光纤，探测双脉冲在传感光纤中产生的背向返回信号经光环行器的第三端口输入第二光耦合器，第二光耦合器将背向返回信号与本地参考光干涉，得到与光调制器移频数值相等的拍频信号，光电探测器将拍频信号转换为电信号，数据采集器对该电信号进行采集和处理，获得传感光纤的扰动信息；	1.位相敏感光時間領域反射計に基づく位相復調方法であって,その特徴は以下のとおりである:該方法は位相復調装置に基づき,前記位相復調装置は狭線幅レーザ,第一光カプラ,光変調器,光増幅器,サーキュレータ,第二光カプラ,センシングファイバ,パルス信号発生器,光検出器及びデータ収集器を含み,前記センシングファイバに弱反射格子が均一に間隔を置いて刻まれ,第二光カプラは50:50光カプラであり,パルス信号発生器は光変調器及びデータ収集器を制御するために用いられる;狭線幅レーザは連続光を発し,該連続光は第一カプラによってパワーが異なる二パス光に分割され,ここでパワーが小さい方の光はローカル参照光とし,パワーが大きい方の光はプローブ光とし,該検出光は光変調器によって検出ダブルパルスに変調され,該検出ダブルパルスは光増幅器で増幅された後に光サーキュレータの第一ポートに入力され,且つ光サーキュレータの第二ポートからセンシングファイバに注入され,検出ダブルパルスによってセンシングファイバで生成されたバックリターン信号は光サーキュレータの第三ポートを経て第二光カプラに入力され,第二光カプラはバックリターン信号をローカル参照光と干渉させ,光変調器の周波数シフト値に等しいビート周波数信号を取得し,光検出器はビート周波数信号を電気信号に変換し,データ収集器は該電気信号に対して収集及び処理を行い,センシングファイバの摂動情報を取得する;	1.位相敏感光時間領域反射計に基づく位相復調方法であって,その特徴は以下のとおりである:該方法は位相復調装置に基づき,前記位相復調装置が狭い線幅レーザ、第一光カプラ、光変調器、光増幅器を含み,サーキュレータ、第2光カプラ、センシングファイバ、パルス信号発生器、光検出器及びデータ収集器を含み,前記センシングファイバには弱い反射格子が均一に間隔をおいて刻まれ,第2光カプラは50:50光カプラであり,パルス信号発生器は光変調器及びデータ収集器を制御するために用いられる;狭線幅レーザは連続光を発し,該連続光は第一カプラによってパワーが異なる2パス光に分割され,ここでパワーが小さい方の光はローカル基準光として,パワーが大きい方の光はプローブ光として,該検出された電気信号を収集及び処理するために,電気信号に基づいて以下のステップを行う:データ収集器が収集した電気信号に基づいて以下のステップを行う:光変調器の周波数変換値と等しいビート周波数信号を取得するために,電力が大きい方の光をプローブ光として,該検出は光変調器が収集した電気信号に基づいて以下のステップを行う:光変調器により収集された電気信号に基づいて以下のステップを行う:電力がより大きい光を検出光とし,該検出光を光変調器により収集した電気信号に基づいて取得する。２３２２。	1. 位相敏感光時間領域反射計に基づく位相復調方法であって、該方法は位相復調装置に基づき,前記位相復調装置は狭線幅レーザ、第一光カプラ、光変調器、光増幅器,サーキュレータ、第二光カプラ、センサファイバ、パルス信号発生器、光電検出器及びデータ収集器を含み,前記センサファイバに弱反射格子が均一に刻まれ,第二光カプラは50:50光カプラであり,パルス信号発生器は光変調器及びデータ収集器を制御するために用いられる;狭線幅レーザは連続光を発し,該連続光は第一カプラによって電力が異なる二つの光に分けられ,そのうち電力が小さい一つの光はローカル参照光とし,電力が大きい一つの光は検出光とし,該検出光は光変調器によって検出デュアルパルスに変調され,該検出デュアルパルスは光増幅器によって増幅された後に光サーキュレータの第一ポートに入力され,且つ光サーキュレータの第二ポートによってセンサファイバに注入され,検出デュアルパルスがセンサファイバにおいて生成した逆方向戻り信号は光サーキュレータの第三ポートによって第二光カプラに入力され,第二光カプラは逆方向戻り信号とローカル参照光を干渉し,光変調器の周波数シフト値に等しいビート信号を取得し,光検出器はビート信号を電気信号に変換し,データ収集器は該電気信号を収集して処理し,センサファイバの摂動情報を取得することを特徴とする。	1.位相敏感光時間領域反射率計に基づく位相復調方法,その特徴は、位相復調装置,前記位相復調装置は、狭線幅レーザと、第1光カプラと、光変調器と、光増幅器と、を備え、サーキュレータ、第二光カプラ、センシング光ファイバ、パルス信号発生器、光検出器及びデータ収集器,前記センシング光ファイバに弱反射格子が均一に間隔を置いて刻まれている,第2の光カプラは50:50 フォトカプラ,光変調器及びデータ収集器を制御するためのパルス信号発生器;狭線幅レーザの連続光放出,該連続光は第一カプラによって電力が等しくない二つの光に分けられる,そのうち電力の小さい方路の光はローカル参照光とする,パワーが大きい方の方の光をプローブ光とする,該プローブ光は光変調器によって検出二重パルスに変調される,該プローブダブルパルスを光増幅器で増幅した後に光サーキュレータの第一ポートに入力する,且つ光サーキュレータの第二ポートからセンシング光ファイバーを注入する,検出デュアルパルスがセンシング光ファイバーで発生したバックターン信号を光サーキュレータの第3ポートを介して第2フォトカプラーに入力する,第2の光カプラは、後方戻り信号をローカル基準光と干渉させる,光変調器の偏移数値に等しいビート周波数信号を得る,光検出器はビート周波数信号を電気信号に変換する,データ収集器は該電気信号を収集して処理する,センシング光ファイバーの摂動情報を取得する;	1.位相敏感光時間領域反射率測定に基づく位相復調方法であって,その特徴は以下のとおりである:該方法は位相復調装置に基づき,前記位相復調装置は狭線幅レーザ、第一光カプラ、光変調器、光増幅器,サーキュレータ、第二光カプラ、センシング光ファイバ、パルス信号発生器、光検出器及びデータ収集器を含み,前記センシング光ファイバに弱反射格子が均一に間隔を置いて刻まれ,第二光カプラは50:50光カプラであり,パルス信号発生器は光変調器及びデータ収集器を制御することに用いられる;狭線幅レーザは連続光を発し,該連続光は第一カプラによってパワーが異なる二つの光に分けられ,そのうちパワーがより小さい一つの光はローカル参照光とし,パワーがより大きい一つの光は検出光とし,該検出光は光変調器によって検出ダブルパルスに変調され,該検出ダブルパルスは光増幅器によって増幅された後に光サーキュレータの第一ポートに入力され,且つ光サーキュレータの第二ポートによってセンシング光ファイバに注入され,検出ダブルパルスがセンシング光ファイバに生成した逆方向リターン信号は光サーキュレータの第三ポートによって第二光カプラに入力され,第二光カプラは逆方向リターン信号とローカル参照光を干渉し,光変調器の周波数シフト値に等しいビート信号を得て,光検出器はビート信号を電気信号に変換し,データ収集器は該	１．位相感応光時間領域反射率計に基づく位相復調方法であって、位相復調装置に基づく方法であって、前記位相復調装置は、狭線幅レーザー、第１の光カプラー、光変調器、光増幅器、サーキュレータ、第２の光カプラ、センシング ファイバ、パルス信号発生器、光検出器、およびデータ コレクタ。ここで、センシング ファイバは弱い反射グレーティングで等間隔に配置され、第２の光カプラは 50:50 の光ファイバです。カプラー、パルス信号発生器は光変調器とデータコレクターを制御するために使用され、狭線幅レーザーは連続光を放出し、連続光は最初のカプラーによって異なるパワーを持つ2つの光路に分割されます。の光路を局部参照光とし、一方のパワーの高い方の光路を検出光とし、検出光を光変調器で変調して検出ダブルパルスとし、検出ダブルパルスを光変調器で増幅する。光サーキュレータの２番目のポートはセンシング ファイバに注入され、センシング ファイバの検出ダブル パルスによって生成されたバック リターン信号は、２番目の光サーキュレータに入力されます。光サーキュレータの３番目のポートを介してカプラーを通過し、２番目の光カプラーが戻りますリターン信号はローカル参照光と干渉して、光変調器の周波数シフト値に等しいビート周波数信号を取得します. 光検出器はビート周波数信号を変換します.データ収集器は、電気信号を収集して処理し、センシングファイバ障害情報を取得します。	機械翻訳プラットフォームで、文の後半の内容が大規模に訳抜けした事例。エンジン①と②は文末まで正しく訳せている（ただしエンジン②はこの文も一節ごとに終止形や体言止めとなっており、日本語文として不自然）。一方、エンジン③は、正しく訳せていたが文末で翻訳文の途絶が発生している。エンジン④は、文末まで訳出自体はされているが、内容は非常に不正確。
	3.3.3. セミコロン区切り文での訳抜け	3.3.3-1	CN106620893B ｾｸｼｮﾝA	セミコロン区切り単位の訳抜け	1.人工晶状体，其包括：	1.眼内レンズであって,それは以下を含む:	1.眼内レンズであって,それは以下を含む:	1. 眼内レンズであって,それは以下を含む:	1.眼内レンズであって,それは以下を含む:	1.眼内レンズであって,それは以下を含む:	1. 眼内レンズであって、	セミコロンで区切られた請求項。機械翻訳プラットフォームでは4文目の内容が訳抜けし、埋め合わせ的に3文目の内容が重複出力されたが、他のエンジンはいずれも問題なく訳出できており、不備は見られない。ただしエンジン③は、最終行に不要な「方法」が補足されている。
					- 基体材料，选自疏水型丙烯酸酯、丙烯酸酯水凝胶、硅胶、硅水凝胶、氟硅丙烯酸酯、聚苯乙烯和聚甲基丙烯酸甲酯、聚碳酸酯、聚硅氧烷，或其混合物；	-マトリックス材料であって,疎水性アクリレート,アクリレートヒドロゲル,シリコーンゲル,シリコーンヒドロゲル,フルオロシリコーンアクリレート,ポリスチレンおよびポリメチルメタクリレート,ポリカーボネート,ポリシロキサン,またはその混合物から選択されるもの;	-疎水性アクリレート,アクリレートヒドロゲル,シリカゲル,シリコーンヒドロゲル,フルオロシリコーンアクリレート,ポリスチレンおよびポリメチルメタクリレート,ポリカーボネート,ポリシロキサンからなる群から選択されるマトリックス材料;	- 基材であって,疎水性アクリレート、アクリレートハイドロゲル、シリカゲル、シリコーンハイドロゲル、フルオロシリコーンアクリレート、ポリスチレン及びポリメチルメタクリレート、ポリカーボネート、ポリシロキサン,又はその混合物から選択されるもの;	-基材であって,疎水性アクリレート、アクリレートヒドロゲル、シリカゲル、シリコーンハイドロゲル、フルオロシリコーンアクリレート、ポリスチレン及びポリメチルメタクリレート、ポリカーボネート、ポリシロキサン,又はその混合物から選択されるもの;	疎水性タイプのアクリレート、アクリレートヒドロゲル、シリカゲル、シリコーンヒドロゲル、フルオロシリコーンアクリレート、ポリスチレンおよびポリメチルメタクリレート、ポリカーボネート、ポリシロキサン、またはそれらの混合物から選択されるマトリックス材料;	－疎水性アクリレート、アクリレートヒドロゲル、シリカゲル、シリコンヒドロゲル、フルオロシリコンアクリレート、ポリスチレンおよびポリメチルメタクリレート、ポリカーボネート、ポリシロキサン、またはそれらの混合物から選択されるマトリックス材料；	
					- 至少一种光敏剂，选自光动力型和/或光热型光敏剂，光敏剂含有可聚合基团；	-少なくとも1種の光増感剤であって,光力学型および/または光熱型から選択され,光増感剤が重合性基を含むもの;	または光力学および/または光熱型光増感剤からなる群から選択される少なくとも1種の光増感剤;	- 少なくとも一種の光増感剤であって,光線力学型及び/又は光熱型光増感剤から選択され,光増感剤が重合性基を含有するもの;	-少なくとも一種の光増感剤であって,光パワー型及び/又は光熱型光増感剤から選択され,光増感剤は重合性基を含有するもの;	光力学的および/または光熱的光増感剤から選択される少なくとも1つの光増感剤であって、重合性基を含有する光増感剤と	・重合性基を含む光力学および／または光熱光増感剤から選択される少なくとも１つの光増感剤；	
					其中，光敏剂与基体材料以共价键方式结合，选自：	ここで,光増感剤はマトリックス材料と共役結合の方式で結合され,以下から選択される:	光活性剤および/または光活性剤からなる群から選択される少なくとも1種の光増感剤;	ここで,光増感剤と基材は共有結合方式で結合され,以下から選択される:	ここで,光増感剤はマトリックス材料と共有結合方式で結合し,以下から選択される:	ここで,光増感剤はマトリックス材料と共有結合し,以下から選択される:	ここで、光増感剤は、共有結合で基材と組み合わされ、以下から選択される：	
					- 光敏剂在基体材料成型过程中参与聚合；和/或	-光増感剤がマトリックス材料の成型過程で重合に加わる;および/または	および/または	- 光増感剤は基材の成形過程において重合に関与する;及び/又は	-光増感剤は基材の成形過程において重合に関与する;及び/又は	光増感剤は、マトリックス材料の成形中に重合に関与する。	- 光増感剤は、マトリックス材料の成形中に重合に関与し、および/または	
					- 光敏剂以表面接枝、表面修饰方式固定在基体材料表面，	-光増感剤が表面グラフト,表面修飾の方式でマトリックス材料表面に固定される;	-光増感剤が表面グラフト化され,表面修飾されて基材表面に固定される。	光増感剤は表面グラフト、表面修飾方式で基材の表面に固定され、	-光増感剤は表面グラフト、表面修飾方式で基材の表面に固定され,	-光増感剤は表面グラフト,表面修飾の方式で基体材料の表面に固定され	- 光増感剤は、表面グラフトおよび表面改質によってマトリックス材料の表面に固定され、	
					其中基于材料总重量，所含光敏剂含量为小于1重量%。	ここで材料の全重量に対して,含まれる光増感剤の含有量は1重量%未満である。	この場合,含有される光増感剤の含有量は,材料の全質量に対して1質量%未満である。	ここで材料の総重量に基づき,含まれる光増感剤の含有量は1重量%未満である。	光増感剤の含有量は、材料の総重量に基づいて1重量%未満である。	前記光増感剤が、前記材料の総重量に基づいて1重量%未満の量で存在する、方法。	材料の総重量に基づいて、光増感剤の含有量は１重量%未満である。	
	3.3.4. 事例集	3.3.4-1	CN106164026B ｾｸｼｮﾝC	鉱物名称羅列での訳抜け	24.用于制备1,1,1-三氟-2,3-二氯丙烷的方法，其包括在以催化有效量的催化剂存在下使氯气与3,3,3-三氟丙烯接触以形成1,1,1-三氟-2,3-二氯丙烷，其中所述催化剂包含至少一种金属卤化物，其中所述金属是来自周期表中的第13、14或15族的金属或过渡金属或其组合，所述氯化反应在液相中进行，其中所述催化剂有效量为存在的反应物的总量的0.1至10重量%，其中所述金属为镍、铬、铁、钪、钇、镧、钛、锆、铪、钒、钼、钨、锰、铼、钌、锇、钴、钯、铜、锌、钽、铝、锡或铅。	24.1,1,1-トリフルオロ-2,3-ジクロロプロパンの製造方法であって,触媒作用有効量の触媒の存在下で塩素ガスを3,3,3-トリフルオロプロペンと接触させて1,1,1-トリフルオロ-2,3-ジクロロプロパンを形成することを含み,ここで前記触媒は少なくとも1種の金属ハロゲン化物を含み,ここで前記金属は周期表の第13族,第14族または第15族からの金属または遷移金属またはそれらの組み合わせであり,前記塩素化反応は気相中で行い,ここで前記触媒の有効量は存在する反応物の総量の0.1～10重量%であり,ここで前記金属はニッケル,クロム,鉄,スカンジウム,イットリウム,ランタン,チタン,ジルコニウム,ハフニウム,バナジウム,モリブデン,タングステン,マンガン,レニウム,ルテニウム,オスミウム,コバルト,パラジウム,銅,亜鉛,タンタル,アルミニウム,スズ又は鉛である。	24.触媒有効量の触媒の存在下で塩素を3,3,3-トリフルオロプロペンと接触させて1,1,1-トリフルオロ-2,3-ジクロロプロパンを形成する工程を含み,前記触媒が少なくとも一つの金属ハロゲン化物を含み,前記金属が周期表の第13族,第14族または第15族の金属または遷移金属またはこれらの組合せであり,前記金属が,存在する反応物質の全量の0.1～10重量%であり,前記金属が,ニッケル,クロム,鉄,スカンジウム,アルミニウム,スズまたは鉛であり,前記触媒が,存在する反応物の総量の～重量%である,1,1,1-トリフルオロ-2,3-ジクロロプロパンの製造方法。	24. 1,1,1-トリフルオロ-2,3-ジクロロプロパンの製造方法であって、1,1,1-トリフルオロ-2,3-ジクロロプロパンを形成するために、触媒有効量の触媒の存在下で塩素ガスを3,3,3-トリフルオロプロペンと接触させることを含み、前記触媒が少なくとも1種の金属ハロゲン化物を含み、前記金属が、周期律表の第13族、第14族、または第15族からの金属または遷移金属、またはそれらの組み合わせであり、前記塩素化反応が液相中で行われ、前記触媒有効量が、存在する反応物の総量の0.1～10重量%であり、前記金属が、ニッケル、クロム、鉄、スカンジウム、イットリウム、ランタン、チタン、ジルコニウム、ハフニウム、バナジウム、モリブデン、タングステン、マンガン、レニウム、ルテニウム、オスミウム、コバルト、パラジウム、銅、亜鉛、タンタル、アルミニウム、スズ、または鉛である、方法。	24.1,1,1-トリフルオロ-2,3-ジクロロプロパンを製造する方法であって,それは触媒有効量の触媒の存在下で塩素ガスを3,3,3-トリフルオロプロペンと接触させて1,1,1-トリフルオロ-2,3-ジクロロプロパンを形成することを含み,ここで前記触媒は少なくとも一種の金属ハロゲン化物を含み,ここで前記金属は周期表における13、14又は15族からの金属又は遷移金属又はその組み合わせであり,前記塩素化反応は液相で行われ,ここで前記触媒有効量は存在する反応物の総量の0.1から10重量%であり,ここで前記金属はニッケル、クロム、鉄、スカンジウム、イットリウム、ランタン、チタン、ジルコニウム、ハフニウム、バナジウム、モリブデン、タングステン、マンガン、レニウム、ルテニウム、オスミウム、コバルト、パラジウム、銅、亜鉛、タンタル、アルミニウム、スズ又は鉛である。	触媒有効量の触媒の存在下で塩素ガスを1,1,1-トリフルオロプロペンと接触させて3,3,3-トリフルオロ-2,3-ジクロロプロパンを形成することを含む1,1,1-トリフルオロ-2,3-ジクロロプロパンの製造方法であって、前記触媒が少なくとも1種の金属ハロゲン化物を含み、前記金属が周期表の第13族、第14族もしくは第15族からの金属もしくは遷移金属またはそれらの組合せであり、前記塩素化反応が液相で行われ、前記触媒有効量が存在する反応物の総量の0.1～10重量%であり、前記金属がニッケル、クロム、鉄、スカンジウム、イットリウム、ランタン、チタン、ジルコニウム、ハフニウム、バナジウム、モリブデン、タングステン、マンガン、レニウム、ルテニウム、オスミウム、コバルト、パラジウム、銅、亜鉛、タンタル、アルミニウム、スズまたは鉛である方法。	触媒的に有効な量の触媒の存在下で塩素ガスを３，３，３－トリフルオロプロペンと接触させて１，１，１－トリフルオロ－２，３－ジクロロプロパンを製造する方法。 1-トリフルオロ-2,3-ジクロロプロパンであって、触媒が少なくとも1つの金属ハロゲン化物を含み、金属が周期表の第13、14、または15族の金属または遷移A金属またはそれらの組み合わせであり、塩素化反応が液相で実施され、触媒は、存在する反応物の総量の０．１から１０重量%の有効量で存在し、金属は、ニッケル、クロム、鉄、スカンジウム、イットリウム、ランタン、チタンであり、ジルコニウム、ハフニウム、バナジウム、モリブデン、タングステン、マンガン、レニウム、ルテニウム、オスミウム、コバルト、パラジウム、銅、亜鉛、タンタル、アルミニウム、スズまたは鉛。	機械翻訳プラットフォームでは、後半の金属名の羅列箇所で大規模な訳抜けが発生した。他のエンジンでは訳抜けは発生せず、全ての金属名称が訳出できている。エンジン④のみ、羅列の途中で「であり」という不要な文言が湧き出しているが、内容理解上はさほど重大な悪影響はない。

他種エンジンによる機械和訳比較結果

報告書　掲載項番		事例番号	採取文献	誤訳の内容	原文	校正文	機械翻訳プラットフォーム	エンジン①	エンジン②	エンジン③	エンジン④	短評
3.4. 湧き出し	3.4.1. 一般的な湧き出し	3.4.1-1	CN105197014B ｾｸｼｮﾝB	訳抜けの補填的な湧き出し①	19.根据权利要求16所述的方法，其中，确定步骤包括：通过融合器基于从行驶车道匹配器和行驶车道跟踪器输出的结果及结果模式确定所述行驶车道匹配结果或者所述行驶车道跟踪结果为所述车辆的所述当前行驶车道。	19.請求項16に記載の方法であって,ここで,判断ステップは以下を含む:融合部を介して,走行車線マッチング部及び走行車線追跡部から出力された結果及び結果パターンに基づいて,前記走行車線のマッチング結果又は前記走行車線の追跡結果が前記車両の現在の走行車線であると判断する。	19.請求項16に記載の方法であって,ここで,決定ステップは以下を含む:前記走行車線マッチング結果又は前記走行車線追跡結果は,前記走行車線マッチング装置及び走行車線追跡装置から出力された結果及び結果パターンに基づいて前記走行車線マッチング結果又は前記走行車線追跡結果が前記車両の現在の走行車線であるか否かを判断するためのものである。	19. 請求項16に記載の方法であって、確定ステップは以下を含む：融合器によって走行車線マッチング器及び走行車線追跡器から出力された結果及び結果モードに基づいて前記走行車線マッチング結果又は前記走行車線追跡結果が前記車両の前記現在の走行車線であると確定する。	19.請求項16に記載の方法であって,ここで,決定ステップは以下を含む:フューマーにより走行車線の整合器及び走行車線の追跡器から出力された結果及び結果モードに基づいて前記走行車線のマッチング結果又は前記走行車線の追跡結果が前記車両の前記現在の走行車線であると決定する。	19.請求項16に記載の方法であって,そのうち,決定ステップは以下を含む:融合装置によって走行車線マッチング装置及び走行車線追跡装置から出力された結果及び結果モードに基づいて前記走行車線マッチング結果又は前記走行車線追跡結果を前記車両の前記現在の走行車線として決定する。	請求項１６に記載の方法であって、前記決定するステップは、レーンマッチャおよびレートントラッカから出力された結果および結果パターンに基づいてフューザによってレーンマッチング結果またはレーントラッキングを決定するステップを含むことを特徴とする方法。車両の。	機械翻訳プラットフォームでは、「通过融合器：融合部を介して」に相当する箇所が訳抜けし、埋め合わせ的に文中の別箇所（前記走行車線マッチング結果又は前記走行車線追跡結果は）が重複出力された。他のエンジンのうち、エンジン①と③が該当箇所を正しく訳せている（融合器によって）。一方、エンジン②の「フューマーにより」は訳語が意味不明。エンジン④は、該当箇所は「フューザによって」とひとまず訳出されているが、文末の「前記車両の現在の走行車線であると判断する」に相当する部分が「車両の。」とほぼ訳抜けしており、翻訳精度は最も低い。
		3.4.1-2	CN112127093B ｾｸｼｮﾝD	訳抜けの補填的な湧き出し②	9.根据权利要求7或8所述的X射线图像采集系统(100)，其中，所述控制单元被配置为基于所述准直器配置来探测所述探测器表面(18)上的准直场(52)的几何结构：	9.請求項7又は8に記載のX線画像収集システム(100)であって,ここで,前記制御ユニットは前記コリメータ構成に基づいて前記検出器表面(18)のコリメーションフィールド(52)の幾何学的構造を検出するように構成される;	9.請求項7又は8に記載のX線画像収集システム(100)であって,ここで,前記制御ユニットは前記コリメータ構成に基づいて前記検出器表面(18)上のコリメートフィールド(52)の幾何学的構造を検出するように配置される;	9. 請求項7又は8に記載のX線画像収集システム(100)であって、制御ユニットは、コリメータ構成に基づいて検出器表面(18)上のコリメート場(52)の幾何学的形状を検出するように構成される。	9.請求項7又は8に記載のX線画像収集システム(100)であって,ここで,前記制御ユニットは前記コリメータ構成に基づいて前記検出器表面(18)におけるコリメート場(52)の幾何学的構造を検出するように構成される;	前記制御ユニットは、前記コリメータ構成に基づいて、前記検出器表面(18)上のコリメートされた視野(52)の幾何学形状を検出するように構成される、請求項7又は8に記載のX線画像取得システム(100)。	請求項７または８に記載のＸ線画像取得システム（１００）であって、前記制御ユニットは、前記視野（５２）の前記コリメータ構成に基づいて、前記検出器表面（１８）上のコリメーションを検出するように構成される、Ｘ線画像取得システム（１００）。	機械翻訳プラットフォームでは、2行目の内容が訳抜けし、3行目の内容が不正確に重複出力されている。一方、他の各エンジンは当該行も正しく訳せている。ただしエンジン④は１～２行目の「几何結构（幾何学的構造）」の訳抜けが発生している。
					其中，所述几何结构包括初级延伸(54)和次级延伸(56)，所述次级延伸横向于所述初级延伸；	ここで,前記幾何学的構造は一次延伸部(54)及び二次延伸部(56)を含み,前記二次延伸部は前記一次延伸部に対して横向きである;	ここで,前記幾何学的構造は一次延長部(54)及び二次延長部(56)を含み,前記二次延長部は前記一次延長部より大きい;	ここで,前記幾何学的構造は一次延長部(54)及び二次延長部(56)を含み,前記二次延長部は前記一次延長部に対して横方向である;	ここで,前記幾何学的構造は一次延伸(54)及び二次延伸(56)を含み,前記二次延伸は前記一次延伸に横方向である;	前記幾何学的構造は、一次延長部(54)と二次延長部(56)とを備え、前記二次延長部は、前記一次延長部を横断し	前記形状は、一次延長部（５４）および二次延長部（５６）を含み、前記二次延長部は、前記一次延長部を横切る。	
					并且所述初级延伸大于所述次级延伸；	且つ前記一次延伸部は前記二次延伸部より大きい;	且つ前記一次延長部より大きい; 前記一次延長部は前記二次延長部より大きい;	且つ前記一次延伸は前記二次延伸より大きい;	且つ前記一次延伸は前記二次延伸より大きい;	前記一次伸長は、前記二次伸長よりも大きく	前記一次伸張は前記二次伸張より大きい；	
					其中，所述探测器表面(18)具有一形状，所述探测器表面的所述形状具有初级探测器延伸和次级探测器延伸，并且所述初级探测器延伸大于所述次级探测器延伸；	ここで,前記検出器表面(18)は形状を有し,前記検出器表面の前記形状は一次検出器延伸部及び二次検出器延伸部を有し,且つ前記一次検出器延伸部は前記二次検出器延伸部より大きい;	ここで,前記検出器表面(18)は形状を有し,前記検出器表面の前記形状は一次検出器延伸及び二次検出器延伸を有し,且つ前記一次検出器延伸は前記二次検出器延伸より大きい;	ここで,前記検出器表面(18)は形状を有し,前記検出器表面の前記形状は一次検出器の延伸及び二次検出器の延伸を有し,且つ前記一次検出器の延伸は前記二次検出器の延伸より大きい;	そのうち,前記検出器表面(18)は形状を有し,前記検出器表面の前記形状は一次検出器の延伸及び二次検出器の延伸を有し,且つ前記一次検出器は前記二次検出器より延伸して延伸する;	前記検出器表面(18)は形状を有し、前記検出器表面の前記形状は、一次検出器範囲および二次検出器範囲を有し、前記一次検出器範囲は、前記二次検出器範囲よりも大きく、	前記検出器表面（１８）は形状を有し、前記検出器表面の前記形状は一次検出器延長部および二次検出器延長部を有し、前記一次検出器延長部は前記二次検出器延長部よりも大きく、	
					并且其中，所述旋转单元被配置为基于所述旋转信号(22)来旋转所述X射线探测器(16)，使得所述准直场(52)被限制约束到所述探测器表面(18)的所述形状。	且つここで,前記回転ユニットは前記回転信号(22)に基づいて前記X線検出器(16)を回転させることで,前記コリメーションフィールド(52)を前記検出器表面(18)の前記形状に制限するように構成される。	またこの場合,回転ユニットは,回転信号(22)に基づいてX線検出器(16)を回転させるように構成されており,その結果,コリメート磁場(52)が検出器面(18)の形状に拘束される。	回転ユニットは、回転信号(22)に基づいてX線検出器(16)を回転させ、コリメート場(52)を検出器表面(18)の形状に拘束するように構成される。	また、前記回転ユニットは、前記回転信号(22)に基づいて前記X線検出器(16)を回転させるように配置され、前記コリメートフィールド(52)が前記検出器表面(18)の前記形状に制限される。	前記回転ユニットは、前記コリメート場(52)が前記検出器表面(18)の前記形状に制約されるように、前記回転信号(22)に基づいて前記X線検出器(16)を回転させるように構成される	回転ユニットは、回転信号（２２）に基づいてＸ線検出器（１６）を回転させ、コリメーション場（５２）が形状の検出器表面（１８）に拘束されるように構成される。	
	3.4.2. 不要な特許用語の湧き出し	3.4.2-1	CN107404922B ｾｸｼｮﾝA	不要な特許用語の湧き出し①	17.包含通过根据前述权利要求中任一项所述的方法获得的乳基蛋白质水解物的组合物。	17.前記請求項のいずれか一項に記載の方法により得られた乳ベースタンパク加水分解物を含む組成物。	17.請求項1～10のいずれか一項に記載の方法により得られた乳ベースタンパク加水分解物を含む組成物。	17. 前述の請求項のいずれか一項に記載の方法によって得られる乳ベースのタンパク質加水分解物を含む組成物。	17.前述の請求項のいずれか一項に記載の方法によって得られる乳ベースタンパク質加水分解物を含む組成物。	請求項1～17のいずれか一項に記載の方法によって得られる乳製品ベースタンパク質加水分解物を含む組成物。	請求項１～１６のいずれか一項に記載の方法によって得られる乳ベースのタンパク質加水分解物を含む組成物。	機械翻訳プラットフォームにおいて、「前記請求項のいずれか一項に記載の」であるべきところ「請求項1～10のいずれか一項に記載の」と、原文にない具体的な請求項番号（しかも誤っている）が湧き出した事例。エンジン①と②は原文どおりに正しく訳出されている。一方、エンジン③は「請求項1～17のいずれか一項に記載」と、機械翻訳プラットフォームと同様の具体的な請求項番号が湧き出す不備が発生している。エンジン④も「請求項１～１６のいずれか一項に記載」と具体的な請求項番号が湧き出しているが、こちらは請求項番号自体は正しいため、不備とは判定しなかった。
		3.4.2-2	CN106995964B ｾｸｼｮﾝD	不要な特許用語の湧き出し②	15.根据权利要求8所述的开口机器(2：1002)，其特征在于，所述间隔件具有上部和下部，并且在上部和下部处沿所述纵向轴线(X10，X1010)成对地重叠。	15.請求項8に記載の開口機械(2;1002)であって,その特徴は以下のとおりである:前記スペーサは上部および下部を有し,且つ上部と下部は前記縦方向軸線(X10,X1010)に沿って対をなすように重なり合う。	15.請求項8に記載の開口機器(2;1002)前記スペーサは上部および下部を有し、前記長手軸に沿って(X10,X1010)対をなして上下に重なり合うように上部および下部であることを特徴とする、請求項1に記載の方法。	15. 請求項8に記載の開口機械(2;1002)であって、スペーサは、上部および下部を有し、上部および下部において長手方向軸(X10、X1010)に沿って対で重なることを特徴とする。	15.請求項8に記載の開口機械(2;1002)であって、前記スペーサは、上部と下部を有し、上部と下部で前記長手軸(X10,X1010)に沿って対になって重なることを特徴とする。	前記スペーサは、上部および下部を有し、前記上部および下部において前記長手方向軸(X10、X1010)に沿って対になって重なり合う、請求項8に記載の開口機械(2;1002)。	前記スペーサが上部および下部を有し、上部および下部で前記縦軸（Ｘ１０、Ｘ１０１０）に沿って対で重なることを特徴とする、請求項８に記載の開口機（２；１００２）。	本例も機械翻訳プラットフォームでは翻訳文末尾に「ことを特徴とする、請求項1に記載の方法」という原文に存在しない文言が湧き出したが、他のエンジンでは発生しなかった。
3.4.3. 事例集		3.4.3-1	CN107406889B ｾｸｼｮﾝC	菌名羅列での訳抜けと湧き出し	18.如权利要求17所述的方法，其中该文库被包含在嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、克劳氏芽孢杆菌、凝结芽孢杆菌、坚硬芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、短小芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌或苏云金芽孢杆菌宿主细胞内。	18.請求項17に記載の方法であって,ここで該ライブラリーはバチルス・アルカロフィラス,バチルス・アミロリケファシエンス,バチルス・プレビス,バチルス・サーキュランス,バチルス・クラウジ,バチルス・コアグランス,バチルス・フィルムス,バチルス・ロータス,バチルス・レンタス,バチルス・リケニフォルミス,バチルス・メガテリウム,バチルス・ブミルス,バチルス・ステアロサーモフィルス,バチルス・サブティリス又はバチルス・チューリンゲンシス宿主細胞内に含まれる。	18.請求項17に記載の方法であって,その特徴は以下のとおりである:該ライブラリーはバチルス・セレウス、バチルス・アミロリケファシエンス、バチルス・セレウス、バチルス・サーキュラ、バチルス・クラウジ、バチルス・セレウス、バチルス・セレウス、バチルス・リケニフォルミス、バチルス・リケニフォルミス、バチルス・リケニフォルミス、バチルス・メガテリウム、バチルス・メガテリウム、バチルス・ブミルス、バチルス・サーモフィルス、バチルスサチリス又はバチルス・チューリンゲンシス宿主細胞内に含まれる。	18. 請求項17に記載の方法であって、このライブラリーは、Bacillus albacillus、Bacillus am ylogenes、Bacillus bacillus、Bacillus cerevis、Bacillus coagulosis、Bacillus cerevis、Bacillus am ylogenes、Bacillus macillus、Bacillus thermophilus、Bacillus subtilis宿主細胞内に含まれる。	18.請求項17に記載の方法であって,ここで該ライブラリーはバチルス・バチルス、バチルス・アミロリケファシエンス、バチルス・プレビリス、環状バチルス・バチルス、バチルス・クラゲス、バチルス・コアグランス、バチルス・ベシラス、バチルス・スブチリス、バチルス・バチルス、バチルス・リケニフォルミス、バチルス・メガテリウム、バチルス・ブミルス、バチルス・ステアロサーモフィラス、バチルス・サブチリス又はバチルス・チューリンゲンシス宿主細胞内に含まれる。	18.請求項17に記載の方法であって,ここで該ライブラリーはバチルス・アルカロフィラス、バチルス・アミロリケファシエンス、バチルス・プレビス、バチルス・サーキュランス、バチルス・クラウジイ、バチルス・コアグランス、バチルス・ステアロサーモフィラス、バチルス・レンタス、バチルス・リケニホルミス、バチルス・メガテリウム、バチルス・ブミルス、バチルス・ステアロサーモフィラス、バチルス・サブチリス又はバチルス・チューリンゲンシス宿主細胞内に含まれる。	前記ライブラリーが、バチルス・アルカロフィルス、バチルス・アミロリケファシエンス、バチルス・プレビス、バチルス・サーキュランス、バチルス・クラウジー、バチルス・コアギュランス、バチルス・ファーマス、バチルス・リケニフォルミス、バチルス・メガテリウム、バチルス・ブミルス、宿主細胞中のバチルス・ステアロサーモフィルス、バチルス・ズブチリスまたはバチルス・チューリンゲンシス。	機械翻訳プラットフォームでは、菌名の羅列部分に重複が多く、訳出不能な名称の訳抜けと埋め合わせ的な湧き出しの結果と推測された。他のエンジンを見ると、エンジン①は全てラテン語で出力されているが、正解文とはあまり合致しておらず、かつ同じ名称が複数回出力されていたり、名称の数が少ない（15種に対し、のべ10種）など、より重度の不備が発生している。エンジン②は正解訳と不正解訳がほぼ半々だが、誤訳個所は所々で「バチルス・バチルス」と、訳出不能箇所に「バチルス」を重複出力しているように見える。エンジン③は、中盤の「バチルス・フィルムス,バチルス・ロータス」の2種を「バチルス・ステアロサーモフィラス」と1種に誤訳（訳抜け？）しているが、それ以外は非常に正確に訳出できている。エンジン④は、「バチルス・ロータス」を「バチルス・カンジダシス」と訳している以外は異表記の範疇と見なせ、こちらも菌名は非常に正確である。ただし「宿主細胞」の係り受けが誤っており、「請求項17に記載の方法」も訳抜けしているなど、他の箇所での不備が多い。

報告書　掲載項番		事例番号	採取文献	誤訳の内容	原文	校正文	機械翻訳プラットフォーム	エンジン①	エンジン②	エンジン③	エンジン④	短評
		3.4.3-2	CN112122652B ｾｸｼｮﾝB	意味不明な文言の湧き出し	4.根据权利要求3所述的掏孔器，其特征在于，驱动件包括拉簧(11)和竖直滑动设置在凹槽(1b)内的驱动杆(12)，拨片(8)位于驱动杆(12)和限位杆(7)之间，拉簧(11)两端分别连接拨片(8)和桶体(1)，驱动杆(12)上设有从上到下依次分布的抵压面一(12b)和限位面一(12a)，限位杆(7)上设有从上到下依次设置的限位面二(7a)和抵压面二(7b)，在拉簧(11)作用下，拨片(8)上端的上下两侧分别抵压在抵压面一(12b)和限位面一(12a)上，拨片(8)下端的上下两侧分别抵压在限位面二(7a)和抵压面二(7b)上，驱动杆(12)上成型有水平设置的压杆(12c)以及设置有使驱动杆(12)具有向下运动趋势的第三弹簧(13)，且压杆(12c)外端伸出凹槽(1b)。	4.請求項3に記載のボアビットであって,その特徴は以下のとおりである:駆動ロッドは引張バネ(11)と凹部(1b)内に垂直に摺動するように設置された駆動ロッド(12)を含み,パドル(8)は駆動ロッド(12)と位置決めロッド(7)との間に位置し,引張バネ(11)の両端はそれぞれパドル(8)とバレル(1)とに接続され,駆動ロッド(12)上には上から下に順次押圧面一(12b)と位置決め面一(12a)が設けられ,位置決めロッド(7)に上から下に順次位置決め面二(7a)と押圧面(7b)が設置され,引張バネ(11)の作用下で,パドル(8)の上端の上下両側はそれぞれ押圧面一(12b)と位置決め面一(12a)に押付けられ,パドル(8)の下端の上下両側はそれぞれ位置決め面二(7a)と押圧面(7b)に押付けられ,駆動ロッド(12)には水平に設置された押圧ロッド(12c)が成形され駆動ロッド(12)に下向きに付勢する第三バネ(13)を有し,且つ押圧ロッド(12c)の外端は凹溝(1b)から突き出る。	4.駆動ロッド(12)と制限レバー(7)との間に位置し,引張バネ(11)2の両端がそれぞれタブ(8)とバレル(1)とに接続され,駆動ロッド(12)上に水平に配置された加圧ロッド(シンノリンc)及び駆動ロッド(タードン)上に水平に設置された加圧ロッド(シンキンc)と,駆動ロッド(タードン)上に水平に配置された加圧ロッド(シンマc)と,駆動ロッド(タードン)上に水平に設置された加圧ロッド(シンマc)と,駆動ロッド(タードン)上に水平に設置された加圧ロッド(シンキンc)と,駆動ロッド(タードン)上に水平に設置された加圧ロッド(シンキンc)と,駆動ロッド(タードン)上に水平に設置された加圧ロッド(シンキンc)と,駆動ロッド(タードン)上に水平に配置された加圧ロッド(シンキンc)と,駆動ロッド(タードン)上に形成されて水平に設置された加圧ロッド12 12 7 2 7 2 3 7 11 8 2 12 12 8 2 2 7 11 2 7 12 12 12 3 13 12 1 1 1	請求項3に記載のタッピングマシンであって、駆動部材は引張バネ(11)及び凹溝(1b)内に垂直に摺動可能に設置される駆動ロッド(12)を含み,パドル(8)は駆動ロッド(12)とストッパロッド(7)との間に位置し,引張バネ(11)の両端はそれぞれパドル(8)及びバレル本体(1)に接続され,駆動ロッド(12)に上から下へ順に分布する第一当接面(12b)及び第一ストッパ面(12a)が設けられ,ストッパロッド(7)に上から下へ順に設置する第二ストッパ面(7a)及び第二当接面(7b)が設けられ,引張バネ(11)の作用で,パドル(8)の上端の上下両側はそれぞれ第一当接面(12b)及び第一ストッパ面(12a)に当接され,パドル(8)の下端の上下両側はそれぞれ第二ストッパ面(7a)及び第二当接面(7b)に当接され,駆動ロッド(12)に水平に設置される押圧ロッド(12c)及び駆動ロッド(12)を下向きに移動させる傾向を有する第三バネ(13)が成形され,且つ押圧ロッド(12c)の外端は凹溝(1b)から延出することを特徴とする。	4.請求項3に記載の穴開け器,その特徴は、駆動部材は引張ばね(11)及び凹溝(1b)内に垂直に摺動して設置される駆動ロッド(12)を含む,パドル(8)は駆動ロッド(12)とリミットロッド(7)の間に位置する,引張バネ(11)の両端はそれぞれパドル(8)とバレル本体(1)に接続される,駆動ロッド(12)に上から下へ順に分布する当接面一(12b)及びストッパ面一(12a)が設けられる,ストッパロッド(7)に上から下へ順に設けられるストッパ面2(7a)と押圧面2(7b)が設けられる,引張バネ(11)の作用で,パドル(8)の上端の上下両側がそれぞれ当接面一(12b)及びストッパ面一(12a)に当接する,パドル(8)の下端の上下両側はそれぞれストッパ面2(7a)と押圧面2(7b)に当接する,駆動ロッド(12)に水平に設置された押圧ロッド(12c)及び駆動ロッド(12)に下向きの運動傾向を有する第三バネ(13)が成形される,且つ押圧ロッド(12c)の外端は凹溝(1b)から延出する。	4.請求項3に記載の穿孔装置であって,その特徴は以下のとおりである:駆動部材は引張バネ(11)及び凹溝(1b)内に垂直に摺動可能に設置される駆動ロッド(12)を含み,パドル(8)は駆動ロッド(12)と制限ロッド(7)との間に位置し,引張バネ(11)の両端はそれぞれパドル(8)及びバレル本体(1)に接続され,駆動ロッド(12)に上から下へ順に分布する第一当接面(12b)及び第一制限面(12a)が設けられ,制限ロッド(7)に上から下へ順に設置される第二制限面(7a)及び第二当接面(7b)が設けられ,引張バネ(11)の作用で,パドル(8)の上端の上下両側はそれぞれ第一当接面(12b)及び第一制限面(12a)に当接し,パドル(8)の下端の上下両側はそれぞれ第二制限面(7a)及び第二当接面(7b)に当接し,駆動ロッド(12)に水平に設置される押圧ロッド(12c)が成形され及び駆動ロッド(12)に下向きの運動傾向を持たせる第三バネ(13)が設置され,且つ押圧ロッド(12c)の外端は凹溝(	【請求項４】 前記駆動部材は、前記溝（１ｂ）内に上下摺動可能に配設された引張バネ（１１）と駆動棒（１２）とからなり、前記ピック（８）は前記溝（１ｂ）の間に位置することを特徴とする請求項３に記載の穴あけ機。駆動ロッド（１２）と制限ロッド（７）、引っ張りばね（１１）の両端がそれぞれピック（８）とバケット本体（１）に接続され、駆動ロッド（１２）が設けられている。押圧面１（１２ｂ）および制限面１（１２ａ）を有する制限ロッド（７）は、上から下に順に配置された制限面２（７ａ）および押圧面２（７ｂ）を備える。引っ張りばね(11)により、ピックピース(8)の上端の上下がそれぞれ押圧面1(12b)と限界面1(12a)に押し付けられ、ピック片(8)の下端をそれぞれ制限面2(7a)と圧力面2(7b)に押し付け、駆動棒(12)は水平に配置された押し棒(12c)で形成され、駆動ロッド（１２）の下方への移動傾向第３のばね（１３）、および押圧ロッド（１２ｃ）の外端は、溝（１ｂ）から突出する。	機械翻訳プラットフォームでは、文の内容が大規模に訳抜けし、特定の文言が何度も繰り返し湧き出しているほか、要素部番が意味不明の語句に置き換わる不備も発生している。 他の各エンジンにおいては、エンジン①は、冒頭の請求項番号が欠落していることを除けば文全体が適切に翻訳されており、高品質である。エンジン②も、一節ごとに終止形／体言止めを用いている点が不自然だが、内容は正確である。一方、エンジン③は文の終盤まで正しく訳出されているが、末尾部分で翻訳文の途絶が発生している。エンジン④は文末まで出力されているが、全体的に翻訳精度が低い。
3.5. 数値の処理	3.5.1. 文末への数値出力	3.5.1-1	CN110723907B ｾｸｼｮﾝC	訳抜け箇所の数字の文末羅列①	7.如权利要求5或6所述的方法，其特征在于，步骤1)中，以8℃/分钟的速率升温至500℃，恒温1小时，以4℃/分钟的速率继续升温至1300℃，并恒温1小时，使原料熔融成液态，将坩埚中的液体倾倒在已预热的模具上固化成型，并迅速放入350℃马弗炉中退火，恒温1小时，随炉降至室温，制得铕掺杂透明玻璃。	7.請求項5又は6に記載の方法であって,その特徴は以下のとおりである:ステップ1)において,8℃/分の速度で500℃まで昇温し,1時間恒温し,4℃/分の速度で引き続き1300℃まで昇温し,且つ1時間恒温し,原料を液状に溶融させ,坩堝中の液体を予熱済みの型に流して硬化成型し,且つ速やかに350℃のマッフル炉に入れて焼鈍し,1時間恒温し,炉に伴って室温まで降温し,ユーロピウムドープ透明ガラスを得る。	7.請求項5又は6に記載の方法であって,その特徴は以下のとおりである:ステップ1)において,8℃/分の速度で500℃まで昇温し,1時間恒温で,4℃/分の速度で1300℃まで昇温し続け,且つ1時間恒温で,原料を液状に溶融させて,ユーロピウムドープ透明ガラスを製造する。350 1。	7. 請求項5または6に記載の方法であって、ステップ1)において,8° C/分の速度で500° Cまで昇温し,1時間恒温し,4° C/分の速度で1300° Cまで昇温し続け,且つ1時間恒温し,原料を液体に溶融させ,坩堝内の液体を予熱された金型に注ぎ込んで硬化成形し,且つ350° Cのマッフル炉に迅速に入れて焼鈍し,1時間恒温し,炉と共に室温まで降温し,ユーロピウムドープ透明ガラスを得ることを特徴とする。	7.請求項5又は6に記載の方法であって,その特徴は以下のとおりである:ステップ1)において,8° C/分の速度で500° Cまで昇温し,1時間恒温し,4° C/分の速度で1300° Cまで昇温し続け,且つ1時間恒温し,原料を溶融して液状にし,るつぼ内の液体を予熱した金型に注入して硬化成形し,且つ迅速に350° Cマッフル炉に入れてアニーリングし,1時間恒温し,炉と共に室温まで冷却し,ユウロピウムドープ透明ガラスを得る。	7.請求項5又は6に記載の方法であって,その特徴は以下のとおりである:ステップ1)において,8℃/分の速度で500℃まで昇温し,1時間恒温し,4℃/分の速度で1300℃まで昇温し続け,且つ1時間恒温し,原料を液体に溶融し,坩堝内の液体を予熱された金型に注いで硬化成形し,且つ350℃のマッフル炉に迅速に入れてアニールし,1時間恒温し,炉と共に室温まで下げ,ユーロピウムドープ透明ガラスを製造する。	工程１）において、温度を８℃／分の速度で５００℃まで上昇させ、温度を１時間一定に保ち、温度を１時間維持することを特徴とする、請求項５または６に記載の方法。 4℃/分の速度で1300℃まで昇温を続け、１時間保持して原料を溶かし液状にし、るつぼ内の液を予熱した金型に流し込み固化成形し、素早く投入する350℃のマッフル炉に入れてアニールし、1時間保持した後、炉内で室温まで下げると、ユーロピウム添加透明ガラスができあがります。	機械翻訳プラットフォームでは、大規模な訳抜けが発生し、その箇所に含まれた数値のみが文末に羅列されている。他のエンジンでは訳抜け自体が発生しておらず、文末への数字の羅列も発生していない。
		3.5.1-2	CN107795872B ｾｸｼｮﾝF	訳抜け箇所の数字の文末羅列②	5.根据权利要求4所述的一种可伸缩型便携灯具，其特征在于活动机构（6）和灯泡座底(5)之间设有伞骨机构（4），所述伞骨机构（4）包括有连接在多个伞骨支柱一（54）上的短伞骨（41），在多个伞骨支柱二（64）上设有长伞骨（42），在长伞骨（42）下端设有伞骨连接点（43），所述短伞骨（41）一端连接伞骨连接点（43），在所述长伞骨（42）上设有多个LED小灯（44）。	5.請求項4に記載の伸縮可能な携帯照明器具であって,その特徴は以下のとおりである:可動機構(6)と電球ベース(5)との間に傘骨機構(4)が設けられ,前記傘骨機構(4)は複数の傘骨柱1(54)に接続される短傘骨(41)を含み,複数の傘骨柱2(64)に長傘骨(42)が設けられ,長傘骨(42)の下端に傘骨接続点(43)が設けられ,前記短傘骨(41)の一端に傘骨接続点(43)が設けられ,前記長傘骨(42)に複数のLED小型ランプ(44)が設けられる。	5.請求項4に記載の伸縮式携帯照明器具であって,その特徴は以下のとおりである:可動機構(6)と電球ベース(5)との間に傘機構(4)が設けられ,前記傘機構(4)は複数の傘支柱(54)に接続される短傘(41)を含み,複数の傘支柱2(64)に長傘(42)が設けられ,長傘(41)の端部には傘骨接続点(43)が設けられ,前記長傘(42)の一端には傘骨接続点()が設けられ,前記長傘骨()には複数のLED小灯(44)が設けられる。42 43。	5. 請求項4に記載の伸縮可能な携帯型照明器具であって、可動機構(6)と電球ベース(5)との間に傘骨機構(4)が設けられ,前記傘骨機構(4)は複数の第一傘骨支柱(54)に接続された短い傘骨(41)を含み,複数の第二傘骨支柱(64)に長い傘骨(42)が設けられ,長い傘骨(42)の下端に傘骨接続点(43)が設けられ,前記短い傘骨(41)の一端は傘骨接続点(43)に接続され,前記長い傘骨(42)に複数のLEDランプ(44)が設けられることを特徴とする。	5.請求項4に記載の伸縮可能な携帯灯具であって,その特徴は以下のとおりである:可動機構(6)と電球ベース(5)との間に傘骨機構(4)が設けられ,前記傘骨機構(4)は複数の傘骨柱1(54)に接続された短親骨(41)を含み,複数の親骨柱2(64)に親骨(42)が設けられ,親骨(42)の下端に親骨接続点(43)が設けられ,前記短親骨(41)の一端は親骨接続点(43)に接続され,前記親骨(42)に複数のLED小ランプ(44)が設けられる。	5.請求項4に記載の伸縮可能な携帯型照明器具であって,その特徴は以下のとおりである:可動機構(6)と電球ソケット底部(5)との間に傘骨機構(4)が設けられ,前記傘骨機構(4)は複数の傘骨支柱一(54)に接続された短い傘骨(41)を含み,複数の傘骨支柱二(64)に長い傘骨(42)が設けられ,長い傘骨(42)の下端に傘骨接続点(43)が設けられ,前記短い傘骨(41)の一端は傘骨接続点(43)に接続され,前記長い傘骨(42)に複数のLEDスモールランプ(44)が設けられる。	前記可動機構（６）と前記電球基部（５）の底部との間に傘リブ機構（４）があり、前記傘リブ機構（４）は、複数の傘骨柱１（５４）上に連結された短傘骨（４１）、複数の傘骨柱２（６４）上に長傘骨（４２）を備え、長傘骨（４２）上に、下端にリブ接続点（４３）を設け、短リブ（４１）の一端をリブ接続点（４３）に接続し、長リブに複数の小型ＬＥＤライト（４４）を配置する（42）。	機械翻訳プラットフォームでは、赤字部分に訳抜けと埋め合わせ的な湧き出しが発生しており、その影響で訳抜け箇所に含まれていた数値が文末に出力されている。エンジン①、②、③は訳抜け自体が発生しておらず、正確に翻訳されている。エンジン④は、訳抜けは生じていないが、一箇所、要素付番が要素から分離している不備がある。
	3.5.2. 漢数字の処理に係る不備	3.5.2-1	CN109736014B ｾｸｼｮﾝD	漢数字から算用数字が派生①	7.根据权利要求6所述的一种六边形FRP筋材网片的制造方法，其特征在于，所述限位块的材质为聚四氟乙烯、聚甲醛或硬质硅胶。	7.請求項6に記載の六角形FRPリブメッシュの製造方法であって,その特徴は以下のとおりである:前記位置決めブロックの材質はポリテトラフルオロエチレン,ポリオキシメチレン又は硬質シリカゲルである。	7.請求項6に記載の6角形FRP筋材メッシュの製造方法であって,その特徴は以下のとおりである:前記ストッパブロックの材質はポリテトラフルオロエチレン,ポリオキシメチレン又は硬質シリカゲルである。4。	7. 請求項6に記載の六角形FRPリブメッシュの製造方法であって、前記ストッパの材質はポリテトラフルオロエチレン、ポリオキシメチレン又は硬質シリカゲルであることを特徴とする。	7.請求項6に記載の六角形FRP筋材メッシュシートの製造方法であって,その特徴は,前記ストッパーブロックの材質がポリテトラフルオロエチレン、ポリオキシメチレン又は硬質シリカゲルであることである。	7.請求項6に記載の六角形FRPリブ材メッシュの製造方法であって,その特徴は,前記ストッパの材質がポリテトラフルオロエチレン、ポリオキシメチレン又は硬質シリカゲルであることである。	【請求項７】 制限ブロックが、ポリテトラフルオロエチレン、ポリオキシメチレンまたは硬質シリカゲルからなる、請求項６に記載の六角形ＦＲＰ補強メッシュの製造方法。	機械翻訳プラットフォームでは、原文中の漢数字を算用数字に置き換える傾向がある。本例も「聚四氟乙烯」から算用数字4が派生して文末に出力されている。他のエンジンにはそのような傾向は見られず、機械翻訳プラットフォーム特有の現象であると見られる。
		3.5.2-2	CN111826764B ｾｸｼｮﾝD	漢数字から算用数字が派生②	S1.称取四氢呋喃与N,N-二甲基甲酰胺混合至均匀，形成混合溶剂，加入所述改性聚丙烯腈，搅拌至均匀后，形成改性聚丙烯腈混合液；	S1.テトラヒドロフランとN,N-ジメチルホルムアミドとを均一に混合し,混合溶媒を形成し,前記改質ポリアクリロニトリルを添加し,均一に撹拌することで,改質ポリアクリロニトリル混合液を形成する;	S4.テトラヒドロフランとN,N-ジメチルホルムアミドとを均一に混合し,混合溶媒を形成し,前記変性ポリアクリロニトリルを加え,均一に撹拌した後,変性ポリアクリロニトリル混合物を形成する	S1.テトラヒドロフランを秤量してN,N-ジメチルホルムアミドと均一に混合し,混合溶媒を形成し,前記変性ポリアクリロニトリルを加え,均一に撹拌した後,変性ポリアクリロニトリル混合液を形成する;	S1.テトラヒドロフランとN,N-ジメチルホルムアミドを秤量して均一に混合し,混合溶媒を形成し,前記変性ポリアクリロニトリルを加え,均一に撹拌した後,改質ポリアクリロニトリル混合液を形成する;	S1.テトラヒドロフランとN,N-ジメチルホルムアミドを均一になるまで混合し,混合溶媒を形成し,前記変性ポリアクリロニトリルを加え,均一になるまで撹拌した後,変性ポリアクリロニトリル混合液を形成する;	S1. テトラヒドロフランと N,N-ジメチルホルムアミドを秤量し、均一になるまで混合して混合溶媒を形成し、変性ポリアクリロニトリルを加え、均一になるまで撹拌して変性ポリアクリロニトリル混合溶液を形成する。	本例では、「四氢呋喃」という化合物名から算用数字4が派生し、かつそれが「S1」の1から置き換わり「S4」と誤訳されている。他のエンジンではそのような傾向は見られない。
		3.5.2-3	CN109044881B ｾｸｼｮﾝA	「两」から算用数字2が派生	4)投入非硫酸盐两性表面活性剂、非硫酸盐非离子表面活性剂，搅拌均匀，降温至40-35℃；出料。	4)非硫酸塩両性界面活性剤,非硫酸塩非イオン界面活性剤を投入し,均一に撹拌し,40-35℃まで降温する;完成品を取り出す。	4)は非硫酸塩性界面活性剤、非硫酸塩系非イオン界面活性剤を投入し,均一に撹拌し,温度を2℃まで下げて40-35℃にする;。	4) 非硫酸塩両性界面活性剤、非硫酸塩非イオン界面活性剤を投入し,均一に撹拌し,40-35° Cまで降温する;排出する。	4)非硫酸塩両性界面活性剤、非硫酸塩非イオン界面活性剤を投入し,均一に撹拌し,40-35° Cに降温する;材料を排出する。	4)非硫酸塩両性界面活性剤、非硫酸塩非イオン界面活性剤を投入し,均一に撹拌し,40-35℃まで降温する;排出する。	４）硫酸塩ではない両性界面活性剤と硫酸塩ではない非イオン性界面活性剤を加え、均一に撹拌し、４０～３５℃に冷却し、排出する。	機械翻訳プラットフォームでは「两」という文字から算用数字2が派生する傾向が強く、本例ではその「2」が原文にない文言を伴って出力され、「2℃まで下げて」という重大な湧き出しが生じている。他のエンジンでは「两」から「2」を派生させる傾向は見られない。
	3.5.3. 漢数字を含む化合物名称における誤訳	3.5.3-1	CN109562045B ｾｸｼｮﾝA	漢数字を含む化学物質名の誤訳	4.根据权利要求1至3中任一项所述的组合物，其中所述组合物还包含由选自锌盐、亚锡盐和氯化十六烷基吡啶的口腔可接受的盐提供的口腔可接受的阳离子型活性剂。	4.請求項1から3のいずれか一項に記載の組成物であって,ここで前記組成物は亜鉛塩,第一スズ塩及び塩化セチルピリジニウムから選択される口腔用として許容可能な塩により提供される口腔用として許容可能なカチオン活性剤をさらに含む。	4.請求項1から3のいずれか一項に記載の組成物であって,ここで前記組成物はさらに亜鉛塩、第一スズ塩及び第16塩化ジアルキルピリジンから選択される口腔に許容可能な塩により提供される口腔内に許容可能なカチオン性活性剤を含む。	4. 請求項1～3のいずれか1項に記載の組成物であって、組成物は、亜鉛塩、第一スズ塩、および塩化セチルピリジンから選択される口腔的に許容される塩によって提供される口腔的に許容されるカチオン性活性剤をさらに含む。	4.前記組成物が、亜鉛塩、第一スズ塩および塩化セチルピリジニウムからなる群から選択される口腔に許容される塩によって提供される口腔に許容されるカチオン性活性剤をさらに含む、請求項1~3のいずれか一項に記載の組成物。	前記組成物が、亜鉛塩、スズ塩、及び塩化セチルピリジニウムから選択される経口的に許容される塩によって提供される、経口的に許容されるカチオン性活性剤を更に含む、請求項1～3のいずれか一項に記載の組成物。	組成物が、亜鉛塩、第一スズ塩および塩化セチルピリジニウムからなる群から選択される経口的に許容される塩をさらに含む、請求項１から３のいずれか一項に記載の組成物。	化合物名に含まれる漢数字が算用数字に置換された結果、化合物名が誤訳される傾向も強い。本例も「氯化十六烷基吡啶」から「16」が派生し、化合物名自体が「第16塩化ジアルキルピリジン」と誤訳された。他のエンジンはいずれも「塩化セチルピリジニウム」又は「塩化セチルピリジン」と正しく訳されている。
	3.5.4. 漢数字「一」の処理	3.5.4-1	CN106962145B ｾｸｼｮﾝA	訳出すべき「一」の欠落	所述流量检测单元（7）包括流量传感器一（71）、流量传感器二（72）、流量传感器三（73）和流量传感器四（74），并且流量传感器一（71）、流量传感器二（72）、流量传感器三（73）和流量传感器四（74）的输出端均与微处理器（17）的输入端连接。	流量測定ユニット(7)は流量センサ1(71),流量センサ2(72),流量センサ3(73)及び流量センサ4(74)を含み,且つ流量センサ1(71),流量センサ2(72),流量センサ3(73)及び流量センサ4(74)の出力端はいずれもマイクロプロセッサ(17)の入力端に接続される。	流量検出ユニット(7)は流量センサ(71)、流量センサ2(72)、流量センサ3(73)及び流量センサ4(74)を含み,且つ流量センサ(71)、流量センサ2(72)、流量センサ3(73)及び流量センサ4(74)の出力端はいずれもマイクロプロセッサ(17)の入力端に接続される。	前記流量検出ユニット(7)は第一流量センサ(71)、第二流量センサ(72)、第三流量センサ(73)及び第四流量センサ(74)を含み,且つ第一流量センサ(71)、第二流量センサ(72)、第三流量センサ(73)及び第四流量センサ(74)の出力端はいずれもマイクロプロセッサ(17)の入力端に接続される。	前記流量検出ユニット(7)は流量センサ1(71)、流量センサ2(72)、流量センサ3(73)及び流量センサ4(74)を含み,且つ流量センサ1(71)、流量センサ2(72)、流量センサ3(73)及び流量センサ4(74)の出力端はいずれもマイクロプロセッサ(17)の入力端に接続される。	前記流量検出ユニット(7)は第一流量センサ(71)、第二流量センサ(72)、第三流量センサ(73)及び第四流量センサ(74)を含み,且つ第一流量センサ(71)、第二流量センサ(72)、第三流量センサ(73)及び第四流量センサ(74)の出力端はいずれもマイクロプロセッサ(17)の入力端に接続される。	流量検出ユニット（７）は、流量センサ１（７１）、流量センサ２（７２）、流量センサ３（７３）および流量センサ４（７４）、ならびに流量センサ１（７１）、流量センサ２（７２）を含む。、流量センサ３（７３）および流量センサ４（７４）の出力端子はすべて、マイクロプロセッサ（１７）の入力端子に接続されている。	中国語では、漢数字「一」は英語の不定冠詞的に使われることもあり、和訳時に訳出されない場合がある。その影響か、機械翻訳プラットフォームでは、訳出すべき漢数字「一」が訳出されないケースが散見された。本例も、１行目では流量传感器、2行目では指示灯にそれぞれ漢数字で「一」～「四」の付番が付随しているが、機械翻訳プラットフォームの翻訳文では「一」のみが訳出されていない。 一方、他のエンジンはいずれも問題なく処理できている。ただしエンジン④は両行とも要素並列の範囲誤認という別種の不備が発生している。
					5.根据权利要求1所述的一种园林灌溉具有精准控制的水体分流装置，其特征在于：所述指示灯（21）包括指示灯一（211）、指示灯二（212）、指示灯三（213）和指示灯四（214），并且指示灯一（211）、指示灯二（212）、指示灯三（213）和指示灯四（214）的输入端均与中央处理器（19）的输出端连接。	5.請求項1に記載の庭園の散水のための正確に制御できる水分流装置であって,その特徴は以下のとおりである:前記表示灯(21)は表示灯1(211),表示灯2(212),表示灯3(213)及び表示灯4(214)を含み,且つ表示灯1(211),表示灯2(212),表示灯3(213)及び表示灯4(214)の入力端はいずれも中央プロセッサ(19)の出力端に接続される。	5.請求項1に記載の庭園灌漑に正確な制御を有する水分流装置であって,その特徴は以下のとおりである:前記表示灯(21)は表示灯(211)、表示灯(2)(212)、表示灯(3)(213)及び表示灯(4)(214)を含み,且つ表示灯(211)、表示灯(2)(212)、表示灯(3)(213)及び表示灯(4)の入力端はいずれも中央処理装置(19)の出力端に接続される。214。	5. 請求項1に記載の庭園灌漑における正確な制御を有する水の分流装置であって、その特徴は以下のとおりである：前記表示ランプ(21)は第一表示ランプ(211)、第二表示ランプ(212)、第三表示ランプ(213)及び第四表示ランプ(214)を含み,且つ第一表示ランプ(211)、第二表示ランプ(212)、第三表示ランプ(213)及び第四表示ランプ(214)の入力端はいずれも中央処理装置(19)の出力端に接続される。	5.請求項1に記載の庭園灌漑の正確な制御を有する水体分流装置であって,その特徴は以下のとおりである:前記表示灯(21)は指示灯一(211)、指示灯二(212)、指示灯三(213)及び指示灯四(214)を含み,且つ指示灯一(211)、指示灯二(212)、指示灯三(213)及び指示灯四(214)の入力端はいずれも中央処理装置(19)の出力端に接続される。	5.請求項1に記載の園芸灌漑用の正確な制御を有する水分流装置であって,その特徴は以下のとおりである:前記指示ランプ(21)は指示ランプ1(211)、指示ランプ2(212)、指示ランプ3(213)及び指示ランプ4(214)を含み,且つ指示ランプ1(211)、指示ランプ2(212)、指示ランプ3(213)及び指示ランプ4(214)の入力端はいずれも中央処理装置(19)の出力端に接続される。	前記表示灯（２１）は、表示灯１（２１１）、表示灯２（２１２）、表示灯３（２１３）を含み、表示灯４(214)、表示灯１(211)、表示灯２(212)、表示灯３(213)、表示灯４(214)の入力端子はすべて中央処理装置 (19) の出力に接続されています。繋がり。	
	3.5.5. 事例集	3.5.5-1	CN107208158B ｾｸｼｮﾝC	請求項番号の複雑な羅列	27.如权利要求1-3、5、8-12、13-14、17-21和23-26中任一项所述的方法，其中所述方法是多重复路的。	27.請求項1-3,5,8-12,13-14,17-21及び23-26のいずれか一項に記載の方法であって,ここで前記方法は多重化して行うことができる。	27.請求項1～23-26のいずれか1項に記載のであって,前記方法は多重反復経路であることを特徴とする方法。1－3,5,8－12,13－14,17－21。	27. 請求項1～3、5、8～12、13～14、17～21および23～26のいずれか1項に記載の方法であって、この方法は、多重反復である。	27.前記方法は、多重反復経路である、請求項1~3、5、8~12、13~14、17~21および23~26のいずれか1項に記載の方法。	前記方法が多重化される、請求項1～3、5、8～12、13～14、17～21、および23～26のいずれか1項に記載の方法。	前記方法が多重化される、請求項１～３、５、８～１２、１３～１４、１７～２１および２３～２６のいずれか一項に記載の方法。	被引用請求項がやや複雑に列挙された事例で、機械翻訳プラットフォームでは正しく処理できず、請求項番号の大多数が文末に除けられている。他のエンジンでは問題なく処理されている。

他種エンジンによる機械和訳比較結果

報告書	掲載項番	事例番号	採取文献	誤訳の内容	原文	校正文	機械翻訳プラットフォーム	エンジン①	エンジン②	エンジン③	エンジン④	短評
		3.5.5-2	CN108697595B ｾｸｼｮﾝA	漢数字から算用数字が派生③	所述非反应性有机溶剂选自丁烷、戊烷、己烷、环己烷、庚烷、癸烷、十六烷、甲苯、二甲苯、乙酸乙酯、乙酸丁酯、甲基乙基酮、甲基异丁基酮、1,4-二噁烷、甲基氯、二氯甲烷、氯仿、四氯化碳中的一种或多种,	前記非反応性有機溶媒はブタン,ペンタン,ヘキサン,シクロヘキサン,ヘプタン,デカン,ヘキサデカン,トルエン,キシレン,酢酸エチル,酢酸ブチル,メチルエチルケトン,メチルイソブチルケトン,1,4-ジオキサン,塩化メチル,塩化メチレン,クロロホルム,四塩化炭素のうちの1種又は複数種から選択され,	前記非反応性有機溶剤はブタン類,ペンタン類,ヘキサン類,シクロヘキサン類,ヘプタン類,デカン類,ジオキサン類,トルエン類,キシレン類,酢酸エチル類,酢酸ブチル類,メチルエチルケトン類,メチルイソブチルケトン類,ジオキサン類,塩化メチル類,塩化メチレン類,クロロホルム類,4塩化炭素のうちの一種又は複数種から選択され,16種又は1,4-2種以上である。	前記非反応性有機溶媒はブタン、ペンタン、ヘキサン、シクロヘキサン、ヘプタン、デカン、ヘキサデカン、トルエン、キシレン、酢酸エチル、酢酸ブチル、メチルエチルケトン、メチルイソブチルケトン、1,4-ジオキサン、塩化メチル、ジクロロメタン、クロロホルム、四塩化炭素のうちの一種又は複数種から選択され、	前記非反応性有機溶媒はブタン、ペンタン、ヘキサン、シクロヘキサン、ヘプタン、デカン、ヘキサデカン、トルエン、キシレン、酢酸エチル、酢酸ブチル、メチルエチルケトン、メチルイソブチルケトン、1,4-ジオキサン、塩化メチル、ジクロロメタン、クロロホルム、四塩化炭素のうちの一種又は複数種から選択され、	前記非反応性有機溶媒はブタン、ペンタン、ヘキサン、シクロヘキサン、ヘプタン、デカン、ヘキサデカン、トルエン、キシレン、酢酸エチル、酢酸ブチル、メチルエチルケトン、メチルイソブチルケトン、1,4-ジオキサン、塩化メチル、ジクロロメタン、クロロホルム、四塩化炭素のうちの一種又は複数種から選択され	非反応性有機溶媒は、ブタン、ペンタン、ヘキサン、シクロヘキサン、ヘプタン、デカン、ヘキサデカン、トルエン、キシレン、酢酸エチル、酢酸ブチル、メチルエチルケトン、メチルイソブチルケトン、1、4－ジオキサンのうちの１つ以上から選択される、塩化メチル、塩化メチレン、クロロホルム、四塩化炭素、	「化合物名に含まれる漢数字から算用数字が派生する」不備と、「派生した算用数字が不要な文言とともに文に組み込まれる」不備とが同時発生した事例。他のエンジンはそもそも算用数字が派生する傾向がなく、このため同様の不備は発生していない。ただしエンジン④では、並列関係にある化合物名の一部がそう扱われておらず、要素並列の範囲誤認が生じている。
		3.5.5-3	CN108193303B ｾｸｼｮﾝD	漢数字から算用数字が派生④	9.根据权利要求8所述的亲水聚酯纤维的制备方法，其特征在于，所述聚醚二醇选自聚乙二醇、聚丙二醇或聚丁二醇中的至少一种。	9.請求項8に記載の親水性ポリエステル繊維の製造方法であって,その特徴は以下のとおりである:前記ポリエーテルグリコールはポリエチレングリコール,ポリプロピレングリコールまたはポリブチレングリコールから選択される少なくとも1種である。	9.請求項8に記載の親水性ポリエステル繊維の製造方法であって,その特徴は以下のとおりである:前記ポリエーテルグリコールはポリエチレングリコール,ポリプロピレングリコールまたはポリブチレングリコールから選択される少なくとも2種である。	9. 請求項8に記載の親水性ポリエステル繊維の製造方法であって、前記ポリエーテルジオールはポリエチレングリコール、ポリプロピレングリコール又はポリテトラメチレングリコールのうちの少なくとも一種から選択されることを特徴とする。	9.請求項8に記載の親水性ポリエステル繊維の製造方法であって,その特徴は,前記ポリエーテルジオールがポリエチレングリコール、ポリプロピレングリコール又はポリブチレングリコールのうちの少なくとも一種から選択されることである。	9.請求項8に記載の親水性ポリエステル繊維の製造方法であって,その特徴は,前記ポリエーテルジオールがポリエチレングリコール、ポリプロピレングリコール又はポリブチレングリコールのうちの少なくとも一種から選択されることである。	【請求項９】 ポリエーテルグリコールが、ポリエチレングリコール、ポリプロピレングリコール、ポリテトラメチレングリコールの少なくとも１種から選ばれることを特徴とする請求項８記載の親水性ポリエステル繊維の製造方法。	本例も「化合物名に含まれる漢数字から算用数字が派生する」不備と「派生した算用数字が文中の別の数値から置き換わる」不備とが同時発生した事例である。他のエンジンでは算用数字が派生する傾向がないため、正しく処理されている。
		3.5.5-4	CN108265104B ｾｸｼｮﾝD	「 S311 」 等の誤訳	2.根据权利要求1所述的构建方法，其特征在于，所述步骤S11包括：步骤S111，将所述组织样品剪碎，得到破碎组织；步骤S112，采用所述红细胞裂解液对所述破碎组织进行裂解，得到所述裂解产物。	2.請求項1に記載の構築方法であって,その特徴は,前記ステップS11が以下を含むことである:ステップS111,前記組織サンプルを切断し,破砕組織を得る;ステップS112,前記赤血球溶解液を用いて前記破砕組織を溶解し,前記溶解産物を得る。	2.請求項1に記載の構築方法であって,その特徴は以下のとおりである:前記ステップS11は以下を含む:ステップS11 11であって,前記組織サンプルを切断して,破砕組織を得る;ステップS11 12,前記赤色細胞可溶化物を用いて前記破砕組織を溶解し,前記溶解産物を得る。	2. 請求項1に記載の構築方法であって、前記ステップS11は以下を含むことを特徴とする：ステップS111,前記組織サンプルを剪断し,破砕組織を得る;ステップS112,前記赤血球溶解液を用いて前記破砕組織を溶解し,前記溶解生成物を得る。	2.請求項1に記載の構築方法であって,その特徴は,前記ステップS11が以下を含むことである:ステップS111,前記組織サンプルを細断し,破砕組織を得る;ステップS112,前記赤血球溶解液を用いて前記破砕組織を溶解し,前記分解生成物を得る。	2.請求項1に記載の構築方法であって,その特徴は,前記ステップS11が以下を含むことである:ステップS111,前記組織サンプルを剪断し,破砕組織を得る;ステップS112,前記赤血球溶解液を用いて前記破砕組織を溶解し,前記分解生成物を得る。	ステップＳ１１が、ステップＳ１１１、前記組織サンプルを小片に切断して、破壊された組織を得るステップと、ステップＳ１１２、前記赤血球溶解物を使用して、前記破壊された組織を破砕するステップと、を含む、請求項１に記載の構築方法。ライセート。	機械翻訳プラットフォームでは、「S111」が「S11 11」、「S112」が「S11 12」となぜか1桁増加して中間スペースが入るという特殊な不備が見られた。他のエンジンでは発生しなかった。
		3.5.5-5	CN107795872B ｾｸｼｮﾝF	「“十”形」の不適訳	1)测量放线：根据图纸中管道走向及标高，用石灰线分别放出管道、管井走向和位置；在管井位置处开挖圆形井坑，在井坑中心处用石灰线撒上“十”形标识，井坑深度应在图纸设计基础上加100mm作为雨水井垫层；	1)測定と墨出し:設計図中の管路の方向及び標高に基づいて,石灰ラインを用いて管路,深井戸の方向及び位置を墨出しする;深井戸の位置で円形ピットを掘削し,ピットの中心で石灰ラインで“十”字形標識を描き,ピットの深さは設計図の設計を基礎として雨水槽被衝層として100mm加算する;	1)測定放出:図面中の管路方向及び標高に基づいて,石灰線を用いてそれぞれ管路、管路の井走向及び位置を放出する;管井戸位置で円形井戸ピットを掘削し,井戸ピットの中心に石灰線で“10”字形標識を描き,井戸ピット深さは図面設計上に100mmを雨井戸マット層として加える;	1) 測定放線:図面における管路の方向及び標高に基づき,石灰線でそれぞれ管路、坑井の方向及び位置を放出する;坑井の位置に円形坑井を掘削し,坑井の中心に石灰線で“十”字形標識を散布し,坑井の深さは図面の設計に基づいて100mmを雨水坑井のクッション層として加えるべきである;	1)測定放線:図面における管路の走向及び標高に基づき,石灰線で管路、管井の走向及び位置をそれぞれ放出する;管井の位置で円形坑井を掘削し,坑井の中心に石灰線で“十”形の標識を散き,坑井の深さは図面の設計の基礎に100mmを加えて雨水井のマット層とする;	1)測定と放出:図面における管路の方向及び標高に基づき,石灰線を用いてそれぞれ管路、管井戸の方向及び位置を放出する;管井戸の位置に円形井戸坑を掘削し,井戸坑の中心に用石の石灰線に“十”字形のマークをつけ,井戸坑の深さは図面の設計に基づいて100 mmを加えて雨水井戸の下地層とする;	1) 測定と設置: 図面のパイプラインの方向と高さに従って、ライン ラインを使用して、パイプラインとチューブ ウェルの方向と位置をそれぞれ解放し、チューブ ウェルの位置に円形のウェル ピットを掘削します。井戸ピットの中央にライムラインで「テン」をまきます。雨水井戸の緩衝材として、図面設計に基づいて井戸ピットの深さに 100 mm を追加する必要があります。	機械翻訳プラットフォームでは漢数字を算用数字に翻訳する傾向があるが、本例は漢数字『十』が象形として用いられており、算用数字にしてしまうと意味をなさなくなる。レアケースではあるが、算用数字への置換にはこのような問題もある。エンジン④も「テン」と訳しており、同様の問題が生じている。その他のエンジンはいずれも漢数字を漢数字のまま反映しているため、本来の意図どおりの翻訳文となっている。
		3.5.5-6	CN111576341B ｾｸｼｮﾝE	「一」の要否判断 ※不備なし	所述沥青输送管(6)上设置有沥青辅助输送机构，该沥青辅助输送机构包括一产生压缩空气的空气压缩机(7)和一三通阀(702)，该三通阀(702)设置在沥青泵(104)和缺损修补罩(4)之间的沥青输送管(6)上，剩下的一个接口与空气压缩机(7)的压缩空气出口连接，以利用压缩空气对沥青输送管(6)内的沥青起到辅助输送的作用；	前記アスファルト搬送管(6)にアスファルト補助搬送機構が設置され,該アスファルト補助搬送機構は圧縮空気を生成する空気圧縮機(7)及び三方弁(702)を含み,該三方弁(702)はアスファルトポンプ(104)と欠陥補修カバー(4)との間のアスファルト輸送管(6)に設置され,残りの一つのポートは空気圧縮機(7)の圧縮空気出口に接続され,圧縮空気でアスファルト搬送管(6)内のアスファルトを補助的に搬送する役割を果たす;	前記アスファルト輸送管(6)にアスファルト補助輸送機構が設置され,該アスファルト補助輸送機構は圧縮空気を生成する空気圧縮機(7)及び1 3方弁(702)を含み,該3方弁(702)はアスファルトポンプ(104)と欠陥補修カバー(4)との間のアスファルト輸送管(6)に設置され,残りの1つのインターフェースは空気圧縮機(7)の圧縮空気出口に接続され,圧縮空気でアスファルト輸送管(6)内のアスファルトを補助輸送する役割を果たす;	前記アスファルト輸送管(6)にアスファルト補助輸送機構が設置され,該アスファルト補助輸送機構は圧縮空気を生成するエアコンプレッサ(7)及び三方弁(702)を含み,該三方弁(702)はアスファルトポンプ(104)と欠陥補修カバー(4)との間のアスファルト輸送管(6)に設置され,残りの一つの接続口はエアコンプレッサ(7)の圧縮空気出口に接続され,それにより圧縮空気を利用してアスファルト輸送管(6)内のアスファルトに対して補助輸送の役割を果たす;	前記アスファルト輸送管(6)にアスファルト補助輸送機構が設置され,該アスファルト補助輸送機構は圧縮空気を生成する空気圧縮機(7)及び三方弁(702)を含み,該三方弁(702)はアスファルトポンプ(104)と欠損補修カバー(4)の間のアスファルト輸送管(6)に設置され,残りの一つのインターフェースは空気圧縮機(7)の圧縮空気出口に接続され,圧縮空気を利用してアスファルト輸送管(6)内のアスファルトを補助輸送の役割を果たす;	前記アスファルト輸送管(6)にアスファルト補助搬送機構が設置され,該三方弁(702)はアスファルトポンプ(104)と欠損修理カバー(4)との間のアスファルト輸送管(6)に設置され,残りの一つのインターフェースはエアコンプレッサ(7)の圧縮空気出口に接続され,圧縮空気を利用してアスファルト輸送管(6)内のアスファルトに対して補助輸送の作用を果たす;	アスファルト搬送管(6)にはアスファルト補助搬送機構が設けられており、アスファルト補助搬送機構は、圧縮空気を発生させるエアコンプレッサ(7)と三方弁(702)と、三方弁(702) が設定されているアスファルト ポンプ (104) と欠陥修理カバー (4) の間のアスファルト デリバリ パイプ (6) で、残りの１つのインターフェイスは、空気圧縮機 (7) の圧縮空気出口に接続され、圧縮された空気を利用します。ピチューメン配送パイプへの空気 (6) 内部のアスファルトは、補助輸送の役割を果たします。	漢数字「一」の取り扱いについて、原文前半の「一三通阀」の「一」は前置詞であり訳出せず、後半の「剩下的一个」は「残りの一つの」という意味なので訳出すべきという難易度の高い事例。機械翻訳プラットフォームは、後者は適切に訳出したが、前者も「一」「三」とも算用数字に置換して「1 3方弁」としてしまっている。 一方、他のエンジンはいずれも前者、後者を正しく取扱い、前者は訳出せず「三方弁」、後者は「残りの一つ（1つ）」と的確に翻訳できている。
		3.5.5-7	CN106996547B ｾｸｼｮﾝF	「一一对应」の誤訳	2.知权利要求1所述的易拆装可调角度的灯具支架，其特征在于，每对所述调节盘中，所述调节盘上的相邻两所述轮齿之间设有定位齿，所述轮齿上设有定位槽，所述定位齿与所述定位槽一一对应配合。	2.請求項1に記載の着脱可能で,角度調整可能な照明器具ホルダであって,その特徴は以下のとおりである:各対の前記調整ディスクにおいて,前記調整ディスク上の隣接する二つの前記歯車の間に位置決め歯が設けられ,前記歯車上に位置決め溝が設けられ,前記位置決め歯が前記位置決め溝に一対一に対応して嵌合される。	2.請求項1に記載の着脱可能な角度調整可能な灯具ホルダであって,その特徴は以下のとおりである:各対の前記調整盤において,前記調整盤における隣接する2つの前記歯車の間に位置決め歯が設けられ,前記歯車座に位置決め溝が設けられ,前記位置決め歯は前記位置決め溝1 1に対応して嵌合される。	2. 請求項1に記載の着脱しやすく角度調整可能なランプホルダであって、各対の前記調整ディスクにおいて,前記調整ディスク上の隣接する二つの前記ギア歯の間に位置決めギアが設けられ,前記ギア歯に位置決め溝が設けられ,前記位置決めギアは前記位置決め溝と一対一に対応して係合することを特徴とする。	2.請求項1に記載の着脱可能な角度調整が容易なランプホルダであって,その特徴は以下のとおりである:各対の前記調整ディスクにおいて,前記調整ディスクにおける隣接する二つの前記ホイール歯の間に位置決め歯が設けられ,前記ホイール歯に位置決め溝が設けられ,前記位置決め歯は前記位置決め溝と一対一で対応して係合する。	2.請求項1に記載の取り外しが容易な角度調整可能な灯具ホルダであって,その特徴は以下のとおりである:各対の前記調整ディスクにおいて,前記調整ディスク上の隣接する二つの前記ギア歯の間に位置決め歯が設けられ,前記ギア歯に位置決め溝が設けられ,前記位置決め歯は前記位置決め溝と一対一に対応して嵌合する。	前記調整ディスクの各対において、位置決め歯が前記調整ディスク上の２つの隣接する前記ギア歯の間に配置され、ギア歯が位置決め溝を備えることを特徴とする、請求項１に記載の容易に分解されかつ調整可能な角度ランプブラケット。 、位置決め歯は位置決め溝と１つずつ一致します。	機械翻訳プラットフォームでは、漢数字を含む頻出慣用表現においても、算用数字に置換する処理が優先された結果、適切に訳せていないことが多い。本例も「一一対応:一対一に対応して」という慣用表現を「1 1に対応して」と不適切に訳している。他のエンジンはエンジン④を除きいずれも「一対一に対応」と問題なく訳出しており、機械翻訳プラットフォームにおける漢数字処理に起因した不備と見なせる。エンジン④の「1つずつ一致します」も、正しい意味は類推可能と判定した。
3.6. 英数字記号の処理	3.6.1. 数式											
		3.6.1-1	CN107847107B ｾｸｼｮﾝH	数式の反映① ※不備なし	9.根据权利要求1所述的医疗用观察装置，其中	9.請求項1に記載の医療用観察装置であって,ここで	9.請求項1に記載の医療用観察装置であって,その特徴は以下のとおりである:	9. 請求項1に記載の医療用観察装置であって、ここで、	9.請求項1に記載の医療用観察装置において、	9.請求項1に記載の医療用観察装置であって,そのうち	請求項１に記載の医用観察装置であって、	文中に含まれる数式を正しく反映できるかが焦点の事例。機械翻訳プラットフォームでは問題なく反映できていたが、他のエンジンも同様に正しく反映できている。
					与所述多个光谱分量对应的所述成像信号是RGB信号，并且	前記複数のスペクトル成分に対応する前記撮像信号はRGB信号であり,且つ	前記複数のスペクトル成分に対応する前記撮像信号はRGB信号であり, 且つ	前記複数のスペクトル成分に対応する前記撮像信号は、RGB信号であり、	前記複数のスペクトル成分に対応する前記撮像信号は、RGB信号であり、	前記複数のスペクトル成分に対応する前記撮像信号は、RGB信号であり	前記複数のスペクトル成分に対応する前記撮像信号はＲＧＢ信号であり、	
					当R、G和B分别是R信号、G信号和B信号的辉度分量，并且r、g和b是基于所述权重并且与所述R信号、所述G信号和所述B信号分别对应的系数时，基于下面列出的公式来计算用于计算所述评估值的评估信号L：	R,G及びBがそれぞれR信号,G信号及びB信号の輝度成分であり,且つr,g及びbが前記重みに基づき且つ前記R信号,前記G信号及び前記B信号にそれぞれ対応する係数である場合,以下に列挙される式に基づいて前記評価値を算出するための評価信号Lを算出する:	R、G及びBがそれぞれR信号、G信号及びB信号の輝度成分であり,且つr、g及びbが前記重みに基づき且つ前記R信号、前記G信号及び前記B信号にそれぞれ対応する係数である場合,以下に列挙される式に基づいて前記評価値を計算するための評価信号Lを計算する:	R、G、Bがそれぞれ R信号、G信号、B信号の輝度成分であり、r、g、bが前記重みに基づいて R信号、G信号、B信号にそれぞれ対応する係数である場合、評価値を算出するための評価信号Lは、以下の式に基づいて算出される:	R、G、B が、それぞれ R信号、G信号、B信号の輝度成分であり、r、g、b が、前記重みに基づいて、前記R信号、前記G信号、および前記B信号にそれぞれ対応する係数であるとき、前記評価値を算出するための評価信号Lは、以下の式に基づいて算出される:	評価値を算出するための評価信号Lは、R、G、Bをそれぞれ R信号、G信号、B信号の輝度成分とし、r、g、b を重みに基づく R信号、G信号、B信号にそれぞれ対応する係数としたとき、次式に基づいて算出される。	R、G、B がそれぞれ R 信号、G 信号、B 信号の輝度成分であり、r、g、b が重みに基づき、R 信号、G 信号に対応する場合、評価値を算出するための評価信号 L は、以下の式に基づいて算出されます。	
					[数学公式1]	[数式1]	[数式1]	[数式1]	[数式1]	[数式1]	【数式１】	
					L＝r×R+g×G+b×B。	L＝r×R+g×G+b×B。	L＝r×R+g×G+b×Bである。	L＝r×R g×G b×B。	L＝r×R+g×G+b×B。	L＝r×R+g×G+b×B。	L＝r×R+g×G+b×B。	
		3.6.1-2	CN112267074B ｾｸｼｮﾝC	数式の反映② ※不備なし	1.一种大功率发动机曲轴用高强韧性贝氏体非调质钢，其特征在于，包括如下重量百分比的化学成分：C：0.32-0.38%，Si：0.63-1.25%，Mn：1.60-1.90%，Cr：0.65-0.85%，Mo：0.08-0.12%，V：0.07-0.15%，Ni：0.10-0.18%，Ti：0.022~0.029%，B：0.0020-0.0030%，Al：0.020~0.030%，P：≤0.010%，S：0.027-0.035%，T.O：≤10ppm，[H]：≤1.0ppm，[N]：65-85ppm，205≤X≤245，其余为Fe和不可避免的杂质元素；	1.大電力エンジンのクランクシャフト用高靭性ベイナイト非調質鋼であって,その特徴は以下のとおりである:以下の重量百分率の化学成分を含む:C:0.32－0.38%,Si:0.63－1.25%,Mn:1.60－1.90%,Cr:0.65－0.85%,Mo:0.08－0.12%,V:0.07－0.15%,Ni:0.10－0.18%,Ti:0.022~0.029%,B:0.0020－0.0030%,Al:0.020~0.030%,P:≤0.010%,S:0.027－0.035%,T.O:≤10wtppm,[H]:≦1.0ppm,[N]:65－85ppm,205≦X≦245,残部はFe及び不可避不純物元素;	1.大電力エンジンのクランクシャフト用高靭性ベイナイト非調質鋼であって,その特徴は以下のとおりである:以下の重量百分率の化学成分を含む:C：0.32－0.38%，Si：0.63－1.25%，Mn：1.60－1.90%，Cr：0.65－0.85%，Mo：0.08－0.12%，V：0.07－0.15%，Ni：0.10－0.18%，Ti：0.022~0.029%，B：0.0020－0.0030%，Al：0.020~0.030%，P:≦0.010%，S:0.027－0.035%，T.O:≦10ppm,[H]:≦1.0ppm，[N]：65－85ppm，205≦X≦245,残部はFe及び不可避不純物元素;	1. ハイパワーエンジンクランク軸用高強靭ベイナイト非調質鋼であって、以下の重量パーセントの化学成分を含むことを特徴とする：C:0.32-0.38%,Si:0.63-1.25%,Mn:1.60-1.90%,Cr:0.65-0.85%,Mo:0.08-0.12%,V:0.07-0.15%,Ni:0.10-0.18%,Ti:0.022～0.029%,B:0.0020-0.0030%,Al:0.020～0.030%,P:≦0.010%,S:0.027-0.035%,T.O:≦10ppm,[H]:≦1.0ppm,[N]:65-85ppm,205≦X≦245,残部はFe及び不可避的不純物元素である;	1.高強靭性ベイナイト非調質鋼であって,その特徴は,以下の重量パーセントの化学成分を含むことである:C:0.32-0.38%,Si:0.63-1.25%,Mn:1.60-1.90%,Cr:0.65-0.85%,Mo:0.08-0.12%,V:0.07-0.15%,Ni:0.10-0.18%,Ti:0.022～0.029%,B:0.0020-0.0030%,Al:0.020～0.030%,P:≦0.010%,S:0.027-0.035%,T.O:≦10ppm,[H]:≦1.0ppm,[N]:65-85ppm,205≦X≦245,残部はFe及び不可避的不純物元素である;	1.高出力エンジンクランクシャフト用の高強靭性ベイナイト非調質鋼であって,その特徴は以下のとおりである:以下の重量百分率の化学成分を含む:C:0.32-0.38%,Si:0.63-1.25%,Mn:1.60-1.90%,Cr:0.65-0.85%,Mo:0.08-0.12%,V:0.07-0.15%,Ni:0.10-0.18%,Ti:0.022～0.029%,B:0.0020-0.0030%,Al:0.020～0.030%,P:≦0.010%,S:0.027-0.035%,T.O:≦10 ppm,[H]:≦1.0 ppm,[N]:65-85ppm,205≦X≦245,残部はFe及び不可避的不純物元素である;	1. 高強度エンジンクランクシャフト用高強度靭性ベイニティック非焼入れ焼戻し鋼であって、重量パーセントで、C:0.32-0.38%、Si:0.63-1.25%、Mn:0.63-1.25%の化学成分を含むことを特徴とする鋼。 ：1.60～1.90%、Cr：0.65～0.85%、Mo：0.08～0.12%、V：0.07～0.15%、Ni：0.10～0.18%、Ti：0.022～0.029%、B：0.0020～0.0030%、Al：0.020-0.030%、P: ≤0.010%、S: 0.027-0.035%、TO: ≤10ppm、[H]: ≤1.0ppm、[N]: 65-85ppm、205≤X≤245、残りは Fe および不可避不純物元素。	本例も長大な数式を含むが、機械翻訳プラットフォーム、他の各エンジンとも、正しく反映できている。
					其中，X=25.4(0.003+0.53C)×(1+0.7Si)×(-1.12+5.1Mn)×(1+2.16Cr)×(1+1.74V)×(1+0.364Ni)×(1+3Mo)，式中各元素所指数值=该元素在钢中含量×100；	ここで,X=25.4(0.003+0.53C)×(1+0.7Si)×(－1.12+5.1Mn)×(1+2.16Cr)×(1+1.74V)×(1+0.364Ni)×(1+3Mo)であり,式中の各元素の数値=該元素の鋼鉄中含有量×100である;	ここで,X=25.4(0.003+0.53C)×(1+0.7Si)×(－1.12+5.1Mn)×(1+2.16Cr)×(1+1.74V)×(1+0.364Ni)×(1+3Mo)であり,式中の各元素の指数値=該元素の鋼鉄中含有量×100である;	ここで,X=25.4(0.003 0.53C)×(1 0.7Si)×(-1.12 5.1Mn)×(1 2.16Cr)×(1 1.74V)×(1 0.364Ni)×(1 3Mo),式における各元素の指数値=該元素の鋼における含有量×100;	ここで,X=25.4(0.003+0.53C)×(1+0.7Si)×(-1.12+5.1Mn)×(1+2.16Cr)×(1+1.74V)×(1+0.364Ni)×(1+3Mo)であり,式中の各元素の指数値=該元素の鋼中の含有量×100である;	ここで,X=25.4(0.003+0.53C)×(1+0.7 Si)×(-1.12+5.1 Mn)×(1+2.16 Cr)×(1+1.74 V)×(1+0.364 Ni)×(1+3 Mo)であり,式中の各元素の指数値=該元素の鋼中の含有量×100である;	このうち、X=25.4(0.003+0.53C)×(1+0.7Si)×(-1.12+5.1Mn)×(1+2.16Cr)×(1+1.74V)×(1+0.364Ni)×(１+ 3Mo)、式中の各元素の指数値 =鋼中の元素の含有量 × 100;	

報告書　掲載項番		事例番号	採取文献	誤訳の内容	原文	校正文	機械翻訳プラットフォーム	エンジン①	エンジン②	エンジン③	エンジン④	短評
3.6.2. 化合物名		3.6.2-1	CN106794157B ｾｸｼｮﾝA	英数字を含む化合物名①	7.如权利要求1所述的液体药物组合物，所述组合物包含：(a)9,10-二甲氧基-2-(2,4,6-三甲基苯基亚氨基)-3-(N-氨基甲酰基-2-氨基乙基)-3,4,6,7-四氢-2H-喹啶并[6,1-a]异喹啉-4-酮或其药学上可接受的盐的颗粒，其浓度为0.01mg/mL～30mg/mL；(b)聚氧乙烯二醇脱水山梨糖醇烷基酯，其浓度为0.1mg/mL～2mg/mL；(c)脱水山梨糖醇烷基酯，其浓度为0.01mg/mL～0.1mg/mL；(d)第一磷酸盐缓冲剂组分，其浓度为5mg/mL～10mg/mL；(e)第二磷酸盐缓冲剂组分，其浓度为5mg/mL～10mg/mL；和(f)张力调节剂，其浓度为2mg/mL～8mg/mL。	7.請求項1に記載の液体医薬組成物であって,前記組成物は以下を含む:(a)9,10-ジメトキシ-2-(2,4,6-トリメチルフェニルイミノ)-3-(N-カルバモイル-2-アミノエチル)-3,4,6,7-テトラヒドロ-2H-ピリミド[6,1-a]イソキノリン-4-オン又はその医薬的に許容される塩の粒子であって,その濃度が0.01mg/mL～30mg/mLであるもの;(b)ポリオキシエチレングリコールソルビタンアルキルエステルであって,その濃度が0.1mg/mL～2mg/mLであるもの;(c)ソルビタンアルキルエステルであって,その濃度が0.01mg/mL～0.1mg/mLであるもの;(d)第一リン酸塩緩衝剤成分であって,その濃度が5mg/mL～10mg/mLであるもの;(e)第二リン酸塩緩衝剤成分であって,その濃度が5mg/mL～10mg/mLであるもの;及び(f)張力調節剤であって,その濃度が2mg/mL～8mg/mLであるもの。	7.請求項1に記載の液体医薬組成物であって,前記組成物は以下を含む:(a)9,10-ジメトキシ-2-(2,4,6-トリメチルフェニル基イミノ)-3-(N-カルバモイル-2-アミノエチル)-3,4,6,7-テトラヒドロ-2H-ピリミド[6,1-a]イソキノリン-4-オンまたはその医薬的に許容される塩の粒子であって,その濃度が0.01mg/mL～30mg/mLであるもの;(b)ポリオキシエチレングリコールソルビタンアルキルエステルであって,その濃度が0.1mg/mL～2mg/mL;(c)ソルビタンアルキルエステルであって,その濃度が5mg/mL～10mg/mLである;及び(d)第一リン酸塩緩衝剤成分であって,その濃度が5mg/mL～mg/mLである;及び5mg/mL～mg/mLである;及び0.01 2 40.1mg/mL;(d)10mg/mL;(e)(f)2mg/mL～8mg/mLの濃度を有する張力調節剤。	7. 請求項1に記載の液体医薬組成物であって、組成物は以下を含む：(a) 9,10-ジメトキシ-2-(2,4,6-トリメチルフェニルイミノ)-3-(N-カルバモイル-2-アミノエチル)-3,4,6,7-テトラヒドロ-2H-ピリミド[6,1-a]イソキノリン-4-オン又はその薬学的に許容される塩の粒子であって,その濃度が0.01mg/mL～30mg/mLであるもの;(b) ポリオキシエチレングリコールソルビタンアルキルエステルであって,その濃度が0.1mg/mL～2mg/mLであるもの;(c) ソルビタンアルキルエステルであって,その濃度が0.01mg/mL～0.1mg/mLであるもの;(d) 第一リン酸塩緩衝剤成分であって,その濃度が5mg/mL～10mg/mLであるもの;(e) 第二リン酸塩緩衝剤成分であって,その濃度が5mg/mL～10mg/mLであるもの;及び(f)張力調整剤であって,その濃度が2mg/mL～8mg/mLであるもの。	7.請求項1に記載の液体医薬組成物,組成物は:(a)9,10-ジメトキシ-2-(2,4,6-トリメチルフェニルイミノ)-3-(N-カルバモイル-2-アミノエチル)-3,4,6,7-テトラヒドロ-2H-ピリミジン[6],1-a]イソキノリン-4-オン又はその薬学的に許容される塩の粒子,その濃度が0.01mg/mL～30mg/mL;(b)ポリオキシエチレングリコールソルビタンアルキルエステルであって,その濃度が0.1mg/mL～2mg/mL;(c)ソルビタンアルキルエステルであって,その濃度が0.01mg/mL～0.1mg/mL;(d)第一リン酸緩衝剤成分であって,その濃度が5mg/mL～10mg/mL;(e)第2リン酸緩衝剤成分,その濃度が5mg/mL～10mg/mL;及び(f)張力調整剤,その濃度は2mg/mL～8mg/mLである。	7.請求項1に記載の液体医薬組成物であって,前記組成物は以下を含む:(a)9,10-ジメトキシ-2-(2,4,6-トリメチルフェニルイミノ)-3-(N-カルバモイル-2-アミノエチル)-3,4,6,7-テトラヒドロ-2H-ピリミド[6,1-a]イソキノリン-4-オンまたはその薬学的に許容される塩の粒子であって,その濃度が0.01mg/mL～30mg/mLであるもの;(b)ポリオキシエチレングリコールソルビタンアルキルエステルであって,その濃度が0.1mg/mL～2mg/mLであるもの;(c)ソルビタンアルキルエステルであって,その濃度が0.01mg/mL～0.1mg/mLであるもの;(d)第一リン酸緩衝剤成分であって,その濃度が5mg/mL～10mg/mLであるもの;(e)第二リン酸緩衝剤成分,その濃度が5mg/mL～10mg/mL;及び(f)張力調整剤,その濃度は2mg/mL～8mg/mLである。	（ａ）９，１０－ジメトキシ－２－（２，４，６－トリメチルフェニルイミノ）－３－（Ｎ－カルバモイル－２－アミノエチル）－３を含む、請求項１に記載の液体医薬組成物。,4,6,7-テトラヒドロ-2H-ピリミド[6,1-a]イソキノリン-4-オンまたはその薬学的に許容される、0.01 mg/mL から 30 mg/mL の濃度の許容される塩の粒子; (b) ポリオキシエチレン０，１ｍｇ／ｍＬから２ｍｇ／ｍＬの濃度のグリコールソルビタンアルキルエステル；（ｃ）脱水ソルビトールアルキルエステル、その濃度は０，０１ｍｇ／ｍＬ～０，１ｍｇ／ｍＬ；（ｄ）第１リン酸緩衝成分、その濃度が５ｍｇ／ｍＬ～１０ｍｇ／ｍＬである；（ｅ）第２リン酸塩が５ｍｇ／ｍＬ～１０ｍｇ／ｍＬの濃度の緩衝成分である；および（ｆ）２ｍｇ／ｍＬ～８ｍｇの濃度の等張化剤である/mL。	英数字記号と漢字とが組み合わされた化合物名称であるが、機械翻訳プラットフォームでは、途中で分断されることなく、一連の化合物名称として正しく扱われている。一箇所誤訳もあるが、漢字部分の誤訳であり、英数字記号と組み合わされたことが原因ではない。エンジン①、エンジン③はこうした誤訳もなく完璧に訳せている。一方、エンジン④は化合物名称が途中で分断される重大な不備が発生している。エンジン②も[6,1-a]が[6],1-a]となる軽微な不備はあるが、こちらは化合物名が分断された結果とは考えにくく、カッコと閉じカッコで数字を括るという学習の副作用ではないかと思われる。ただしエンジン②は、入力文が長大だと一節ごとに終止形となるという本エンジン特有の不備がここでも発生している。
		3.6.2-2	CN113622052B ｾｸｼｮﾝD	英数字を含む化合物名②	7.根据权利要求6所述的阻燃单丝的制备方法，其特征在于，所述第二阻燃剂为乙基磷酸二乙酯：所述柔软剂选自间苯二酚双(磷酸二苯酯)和双酚A-双(磷酸二苯酯)中的至少一种；所述第二交联剂位合环氧基团的多功能聚合物：所述抗氧剂为质量比为2：3的β-(3，5-二叔丁基-4-羟基苯基)丙酸正十八碳酯酯和双(2，4-(二苯代)叔丁基苯基)季戊四醇撑双亚磷酸酯的混合物。	7.請求項6に記載の難燃性モノフィラメントの製造方法であって,その特徴は以下のとおりである:前記第二難燃剤はエチルホスホン酸ジエチルである;前記柔軟剤はレゾルシノールビス(ジフェニルホスフェート)及びビスフェノールA-ビス(ジフェニルホスフェート)から選択される少なくとも1種である;前記第二架橋剤はエポキシ基含有多機能ポリマーを含む;前記難化防止剤は質量比が2:1である2 23のβ-(3,5-2-tert-ブチル-4-ヒドロキシフェニル)プロピオン酸n-ヘキサノールと ビス(2,4-(ジフェニル)tert-ブチルフェニル)ペンタエリスリトールビスリン酸エステルの混合物。18 2。	7.請求項6に記載の難燃性モノフィラメントの製造方法であって,その特徴は以下のとおりである:前記第二難燃剤はエチルリン酸ジエチルである;前記柔軟剤はレゾルシノールビス(ジフェニルホスフェート)及びビスフェノールA-ビス(ジフェニルホスフェート)からなる群から選択される少なくとも1種である;前記第二架橋剤はエポキシ基含有多機能ポリマーを含む;前記難化防止剤は質量比が2:1である2 23のβ-(3,5-2-tert-ブチル-4-ヒドロキシフェニル)プロピオン酸n-ヘキサノールと ビス(2,4-(ジフェニル)tert-ブチルフェニル)ペンタエリスリトールビスリン酸エステルの混合物であることを特徴とする。	7. 請求項6に記載の難燃性モノフィラメントの製造方法であって、前記第二難燃剤はエチルリン酸ジエチルである;前記柔軟剤はレゾルシノールビス(リン酸ジフェニル)及びビスフェノールA-ビス(リン酸ジフェニル)から選択される少なくとも一種である;前記第二架橋剤はエポキシ基を含む多機能ポリマー;前記難化防止剤は質量比が2:3のβ-(3,5-ジ-tert-ブチル-4-ヒドロキシフェニル)プロピオン酸n-オクタデカノールエステルと ビス(2,4-(ジフェニル)-tert-ブチルフェニル)ペンタエリスリトールビスホスファイトの混合物であることを特徴とする。	7.請求項6に記載の難燃性モノフィラメントの製造方法であって,その特徴は以下のとおりである:前記第二難燃剤はエチルリン酸ジエチルである;前記柔軟剤はレゾルシノールビス(リン酸ジフェニル)及びビスフェノールA-ビス(リン酸ジフェニル)のうちの少なくとも一種から選択される;前記第二架橋剤はエポキシ基を含有する多機能ポリマーである;前記難化防止剤は質量比が2:3のβ-(3,5-ジ-tert-ブチル-4-ヒドロキシフェニル)プロピオン酸n-オクタデシルエステルと ビス(2,4-(ジフェニル)tert-ブチルフェニル)ペンタエリスリトールビスホスファイトの混合物である。	7.請求項6に記載の難燃性モノフィラメントの製造方法であって,その特徴は以下のとおりである:前記第二難燃剤はエチルリン酸ジエチルである;前記柔軟剤はレゾルシノールビス(ジフェニルホスフェート)及びビスフェノールA-ビス(ジフェニルホスフェート)のうちの少なくとも一種から選択される;前記第二架橋剤はエポキシ基を含有する多機能ポリマーである;前記難化防止剤は質量比が2:3のβ-(3,5-ジ-tert-ブチル-4-ヒドロキシフェニル)プロピオン酸n-オクタデシルエステルと ビス(2,4-(ジフェニル)tert-ブチルフェニル)ペンタエリスリトールビスホスファイトと の混合物である。	請求項６に記載の難燃性モノフィラメントの製造方法は、前記第２の難燃剤がジエチルエチルホスフェート：フェニルエステル）およびビスフェノールＡ－ビス（ジフェニルホスフェート）であり：第２の架橋剤の位置が多官能性である、ことを特徴とする。エポキシ基を含むポリマー；酸化防止剤は質量比２：３-(3,5-ジ-tert-ブチル-4-ヒドロキシフェニル) n-オクタデシル プロピオネートとビス(2,4-(ジフェニル) tert-ブチルフェニル) ペンタエリスリトールbis ホスファイトの混合物。	本例にも英数字記号と漢字が組み合わされた化合物名が3種記載されている。機械翻訳プラットフォームの翻訳又は、このうち1種のみ適訳で、残り2種は一部に誤りがあるが、少なくとも英数字・漢字混在の化合物名をその単位で扱えており、文字種の違いで分割されたりはしていない。他のエンジンでも状況は全く同様であった。なお、機械翻訳プラットフォームでは化合物名に含まれる漢数字から派生した算用数字が文末に「18 2」と羅列されているが、同様の現象は他のエンジンでは見られない。
3.6.3. 事例集		3.6.3-1	CN109667023B ｾｸｼｮﾝD	中国語を含む数式	（5）接結組织在面料的表面形成方格形图案，方格的边长由一花的经、纬向循环数决定，经纬向循环数越大，方格也就越大； 本组织设计时，方格和边长设定为2.5cm，2个方框为一个完整的经向循环； 表层经纱的密度为256根/10cm，表层经纱的循环数=256根/10cm*5cm/10=128根，表层组织为平纹，因此一花循环为64个循环； 表、里层一花总的循环数为256根； 纬纱的排列比为1:1:1，每层的纬纱循环数为6纬，表层纬密为228根/10cm，为使表层形成方格图案，表层组织的纬纱的循环数=表层纬密/10*纬向一花宽度=22.8根/cm*2.5cm=57根纱； 表层组织的纬纱循环数为2，因此表层组织的循环个数为28或29个循环，然后加一根接结纬； 本面料选用28个循环+一根接结纬，纬向基础组织循环数为6纬，纬向总循环数为28*6+一根接结纬，共计169纬；	(5)接合組織は生地の表面に格子状パターンを形成し,菱盤目の辺長は花の経,緯方向のループ数によって決定され,経,緯方向のループ数が大きいほど,格子も大きくなる; 本組織の設計時に,格子と辺の長さは2.5cmに設定され,2つのブロックは一つの完全な経方向ループである; 表層経糸の密度は256本/10cmであり,表層経糸のループ数=256本/10cm*5cm/10=128本であり,表層組織は平織であるため,一花サイクルは64サイクルである; 表、裏層の一花総サイクル数は256本である; 緯糸の配列比が1:1:1であり,各層の緯糸ループ数が6ピックであり,表層の緯密が228本/10cmであり,表層に格子状パターンを形成させ,表層組織の緯糸のループ数=表層の緯密/10*緯方向一花の幅=22.8本/cm*2.5cm=57本糸である; 表層組織の緯糸循環数は2であり,従って表層組織のループ個数は28又は29個ループであり,その後に一本の接合緯入れを行う; 本生地は28個のループ+1本の結合緯入れを選択し,緯方向基礎組織ループ数は6緯であり,緯方向総ループ数は28*6+1本の接合緯入れであり,合計169緯である;	(5)接合組織は生地の表面に格子状パターンを形成し,菱盤目の辺長は花の経,緯周期の循環数によって決定され,緯周期循環数が大きいほど,格子も大きくなる; 本組織の設計時に,格子と辺の長さは2.5cmに設定され,2つのブロックは一つの完全な経方向循環である; 表層経糸の密度は256本/10cmであり,表層経糸の循環数=256本/10cm*5cm/10=128本であり,表層組織は平織であるため,一花サイクルは64サイクルである; 表、裏層の総サイクル数は256本である; 緯糸の配列比が1:1:1であり,各層の緯糸サイクル数が6ピックであり,表層の横糸が228本/10cmであり,表層に格子パターンを形成させ,表層組織の緯糸のループ数=表層の緯密/10*緯方向一花の幅=22.8本/cm*2.5cm=57本糸である; 表層組織の横糸循環数は2であり,従って表層組織の循環数は28又は29サイクルであり,その後に一つの接合緯入れを行う;cm*2.5cm=57 本生地は28つの循環+1本の結合緯入れを選択し,緯側基礎組織循環数は6本であり,緯側総循環数は28*6+1本の接合緯入れであり,合計169本である;	(5) 接合組織は生地の表面に格子状パターンを形成し,格子の一辺の長さは経、緯方向の循環数によって決定され,経緯方向の循環数が大きいほど,格子も大きくなる; 本組織を設計する時,格子及び辺長を2.5cmに設定し,2つのブロックは完全な経方向循環である; 表層経糸の密度は256本/10cmであり,表層経糸のサイクル数=256本/10cm*5cm/10=128本であり,表層組織は平織であり,従って一回のサイクルは64サイクルである; 表、裏層の総サイクル数は256本である; 緯糸の配列比は1:1:1であり,各層の緯糸循環数は6緯であり,表層緯密は228本/10cmであり,表層に格子パターンを形成するために,表層組織の緯糸の循環数=表層緯密/10*緯方向一花幅=22.8本/cm*2.5cm=57本の糸である; 表層組織の緯糸循環数は2であるため,表層組織の循環数は28又は29サイクルであり,続いて一本の接結緯を加える;	(5)接合組織は生地の表面に市松模様を形成し,格子の辺長は花の経緯サイクル数によって決定され,経緯方向の循環数が大きければ大きいほど,マス目も大きくなる; 本組織の設計時に,格子と辺の長さは2.5cmに設定され,2つのブロックは完全な経方向循環である; 表層経糸の密度は256本/10cmであり,表層経糸のサイクル数=256本/10 cm*5cm/10=128本であり,表層組織は平織であり,従って1つの花のサイクルは64サイクルである; 表、裏層一花の総循環数は256本である; 横糸の配列比は1:1:1であり,各層の横糸の繰り返し数は6ピックであり,表層ピック数は228本/10 cmであり,表層に市松模様を形成させるために,表層組織の横糸の繰り返し数=表層ピック数/10*横糸方向の1つの花の幅=22.8本/cm*2.5 cm=57本の糸である; 表層組織の緯糸サイクル数は2であり,従って表層組織のサイクル数は28又は29サイクルであり,次に1本の接結緯糸を加える;	(5)結合組織は生地の表面に格子状パターンを形成し,格子の辺の長さは1つの花の経、緯方向のサイクル数によって決定され,経緯方向のサイクル数が大きいほど,格子も大きくなる; 本組織を設計する時,格子及び辺長を2.5 cmに設定し,2つのブロックは1つの完全な経方向循環である; 表層経糸の密度は256本/10 cmであり,表層経糸のサイクル数=256本/10 cm*5cm/10=128本であり,表層組織は平織であり,従って1つの花のサイクルは64サイクルである; 表面経糸の密度は256/10cm、表面経糸のサイクル数=256/10cm*5cm/10=128、表面織りは平織りなので、1花周期は64サイクルです。 １つの花のループの数は、表面と内側の層の合計で256です。 よこ糸の配置比率は1:1:1、各層のよこ糸周期数は6本、表層のよこ糸密度は228本/10cm 10×よこ方向の花1本分の幅=22.8 ヤーン/cm*2.5cm=57 ヤーン; 表面織りのよこ糸サイクル数は２なので、表面織りのサイクル数は２８または２９サイクルであり、次に結合横糸を追加する。 この生地は28サイクル+結び緯糸1本を使用し、緯糸基本織りサイクル数は6緯糸であり、緯糸サイクルの総数は28*6 +結び緯糸1本、合計169本の緯糸です。	3行目、5行目、7行目に漢字混じりの数式が存在する。機械翻訳プラットフォームでは5行目が一連の数式として取り扱われず、末尾部分が分離されて6行目に出力されている。エンジン④も、5行目で数式を含む文の後半部分に大きな欠落があり、7行目の数式も不完全である。その他の各エンジンは全ての数式を正しい単位で扱えているが、唯一、エンジン③で「28*6+1本」であるべきところ「28*6 一本」と、「+」が訳抜けした。
		3.6.3-2	CN105441384B ｾｸｼｮﾝC	英数字文字列の長大な羅列 ※不備なし	2.权利要求1所述的方法，其中所述MAPK通路抑制因子包括AEG 3482,BI 78D3,CEP1347,c-JUN peptide,IQ 3,JIP-1(153-163),SP 600125,SU 3327,TCS JNK 5a,TCS JNK6o,AMG 548,CMPD-1,EO 1428,JX 401,ML 3403,RWJ 67657,SB 202190,SB 203580,SB203580hydrochloride,SB 239063,SCIO 469hydrochloride,SKF 86002dihydrochloride,SX 011,TAK 715,VX 702,VX 745,Arctigenin,10Z-Hymenialdisine,PD 0325901,PD184352,PD 198306,PD 334581,PD 98059,SL 327,U0124和U0126中的任意一种或多种。	2.請求項1に記載の方法であって,ここで前記MAPK経路抑制因子はAEG 3482,BI 78D3,CEP1347,c−JUN peptide,IQ 3,JIP-1(153-163),SP 600125,SU 3327,TCS JNK 5a,TCS JNK6o,AMG 548,CMPD−1,EO 1428,JX 401,ML 3403,RWJ 67657,SB 202190,SB 203580,SB203580hydrochloride,SB 239063,SCIO 469hydrochloride,SKF 86002dihydrochloride,SX 011,TAK 715,VX 702,VX 745,Arctigenin,10Z−Hymenialdisine,PD 0325901,PD184352,PD 198306,PD 334581,PD 98059,SL 327,U0124及びU0126のいずれか一つ又は複数を含む。	2.請求項1に記載の方法であって,ここで前記MAPK経路抑制因子はAEG 3482,BI 78D3,CEP1347,c−JUN peptide,IQ 3,JIP−1(153−163),SP 600125,SU 3327,TCS JNK 5a,TCS JNK6o,AMG 548,CMPD−1,EO 1428,JX 401,ML 3403,RWJ 67657,SB 202190,SB 203580,SB203580hydrochloride,SB 239063,SCIO 469hydrochloride,SKF 86002dihydrochloride,SX 011,TAK 715,VX 702,VX 745,Arctigenin,10Z−Hymenialdisine,PD 0325901,PD184352,PD 198306,PD 334581,PD 98059,SL 327,U0124及びU01 26のいずれか1つ又は複数を含む。	2. 請求項1に記載の方法であって、前記MAPK経路阻害因子はAEG 3482を含む,BI 78D3,CEP1347,c-JUN peptide,IQ 3,JIP-1(153-163),SP 600125,SU 3327,TCS JNK 5a,TCS JNK6o,AMG 548,CMPD-1,EO 1428,JX 401,ML 3403,RWJ 67657,SB 202190,SB 203580,SB203580 hydrochloride,SB 239063,SCIO 469hydrochloride,SKF 86002dihydrochloride,SX 011,TAK 715,VX 702,VX 745,Arctigenin,10Z-Hymenialdisine,PD 0325901,PD184352,PD 198306,PD 334581,PD 98059,SL 327,U0124及びU0126のうちの任意の一種又は複数種を含む。	2.請求項1に記載の方法,前記MAPK経路阻害因子はAEG 3482を含む,BI 78D3,CEP1347,c-JUN peptide,IQ 3,JIP-1(153-163),SP 600125,SU 3327,TCS JNK 5a,TCS JNK6o,AMG 548,CMPD-1,EO 1428,JX 401,ML 3403,RWJ 67657,SB 202190,SB 203580,SB203580hydrochloride,SB 239063,SCIO 469hydrochloride,SKF 86002dihydrochloride,SX 011,TAK 715,VX 702,VX 745,Arctigenin,10Z-Hymenialdisine,PD 0325901,PD184352,PD 198306,PD 334581,PD 98059,SL 327,U0124およびU0126のいずれか1つまたは複数を含む、請求項1に記載の方法。	MAPK経路阻害剤が、AEG 3482、BI 78D3、CEP1347、c-JUNペプチド、IQ 3、JIP-1 (153-163)、SP 600125、SU 3327、TCS JNK 5a、TCSを含む、請求項1に記載の方法。JNK6o、AMG 548、CMPD-1、EO 1428、JX 401、ML 3403、RWJ 67657、SB 202190、SB 203580、SB203580塩酸塩、SB 239063、SCIO 469塩酸塩、SKF 86002二塩酸塩、SX 011、TAK 715、VX 702、VX 745、Arctigenin、10Z-Hymenialdisine、PD 0325901、PD184352、PD 198306、PD 334581、PD 98059、SL 327、U0124、およびU0126のいずれか1つまたは複数を含む、請求項1に記載の方法。	文中に長大な英数字文字列が存在する事例だが、機械翻訳プラットフォームの翻訳文では正確に転記されている。他のエンジンもほぼ同様だが、エンジン④のみ、文末に置かれるべき「を含む、請求項1に記載の方法。」という文言が英数字文字列の途中の中端や位置に出力されている。なお、エンジン③では、peptideやtydrochlorideを「ペプチド」「塩酸塩」と和訳し、ArctigeninやHymenialdisineは英文字のままと中途半端であるが、内容理解上は問題ないため正解とした。
3.7. 請求項の記載様式に係る不備	3.7.1. 請求項番号が本文の一部として扱われる不備	3.7.1-1	CN108180408B ｾｸｼｮﾝF	請求項番号の誤解釈①	1.一种中继镜头，其特征在于，所述中继镜头包括第一部分和第二部分；所述第一部分的第一入射面为与光路垂直的平面；所述第一部分的第一出射面为凸球面；所述第二部分位于所述第一部分的中心位置；所述第二部分的第二入射面为与所述光路垂直的平面；所述第二部分的第二出射面的曲率小于所述第一部分的第一出射面的曲率；所述第二出射面的最大截面尺寸，通过如下方式得出：计算不同的所述第二出射面的截面尺寸与照明效率及照明均匀性的对应关系；选择所述照明效率和照明均匀性中的任一项达到设定最低值时对应的尺寸，作为所述第二出射面的最大截面尺寸。	1.リレーレンズであって,その特徴は以下のとおりである:前記リレーレンズは第一部分と第二部分とを具備し,前記第一の部分の第一の入射面は,光路と垂直な平面であり,前記第一の部分の第一の出射面が凸球面である前記第二の部分は前記第一部分の中心位置に位置しており,前記第二の部分の第二の入射面が,前記光路と垂直な平面であり,前記第二の部分の第二の出射面の曲率が前記第一の部分の第一の出射面の曲率よりも小さく,前記第二の出射面の最大断面寸法は,前記第二出射面の異なる断面サイズと光効率と照明均一度との対応関係を算出し,前記照明効率および照度の均一性のいずれかに設定され,最も低い値に対応する大きさに到達するように選択し,前記第二の出射面の最大断面寸法であることを特徴とするリレーレンズである。	請求項1に記載の方法において,前記リレーレンズを第一部分と第二部分とを具備し,前記第一の部分の第一の入射面は,光路と垂直な平面であり,前記第一の部分の第一の出射面が凸球面である前記第二の部分は前記第一部分の中心位置に位置しており,前記第二の部分の第二の入射面が,前記光路と垂直な平面であり,前記第二の部分の第二の出射面の曲率が前記第一の部分の第一の出射面の曲率よりも小さく,前記第二の出射面の最大断面寸法は,以下の方式によって得られる:前記第二出射面の断面サイズごとの照明効率と照明均一性との対応関係を算出する;前記照明効率および照明均一性のいずれか一つが設定された最低値に達した時の対応する寸法を選択し,前記第二出射面の最大断面寸法とする。	1. リレーレンズであって、前記リレーレンズは第一部分及び第二部分を含む;前記第1部分の第1入射面は、光路に垂直な平面であり;前記第一部分の第一出射面は凸球面である;前記第二部分は前記第一部分の中心位置に位置する;前記第2部分の第2入射面は、前記光路に垂直な平面であり;前記第二部分の第二出射面の曲率は前記第一部分の第一出射面の曲率より小さく;前記第2出射面の最大断面寸法であることを特徴とする：異なる前記第二出射面の断面寸法と照明効率及び照明均一性との対応関係を計算する;前記照明効率および照明均一性のいずれかが設定最低値に達した時に対応する寸法を選択し,前記第二出射面の最大断面寸法とする。	1.リレーレンズであって,その特徴は以下のとおりである:前記リレーレンズは第一部分及び第二部分を含む;前記第一部分の第一入射面は光路に垂直な平面である;前記第一部分の第一出射面は凸球面である;前記第二部分は前記第一部分の中心位置に位置する;前記第二部分の第二入射面は前記光路に垂直な平面である;前記第二部分の第二出射面の曲率は前記第一部分の第一出射面の曲率より小さい;前記第二出射面の最大断面寸法は,以下の方式により得られる:異なる前記第二出射面の断面寸法と照明効率及び照明均一性との対応関係を計算する;前記照明効率および照明均一性のいずれか一項が設定された最低値に達する時に対応する寸法を選択し,前記第二出射面の最大断面寸法とする。	1.リレーレンズであって,その特徴は以下のとおりである:前記リレーレンズは第一部分及び第二部分を含む;前記第一部分の第一入射面は光路に垂直な平面である;前記第一部分の第一出射面は凸球面である;前記第二部分は前記第一部分の中心位置に位置する;前記第二部分の第二入射面は前記光路に垂直な平面である;前記第二部分の第二出射面の曲率は前記第一部分の第一出射面の曲率より小さい;前記第二出射面の最大断面寸法は,以下の方式によって得られる:異なる前記第二出射面の断面寸法と照明効率及び照明均一性との対応関係を計算することにより、第２出射面の最大断面サイズが得られます; 対応するサイズを選択します発光効率と発光均一度のいずれかが設定最小値に達したとき、第２出射面の最大断面サイズとして設定されます。	1. リレー レンズは、第１部分と第２部分を含み、第１部分の第１入射面は光路に垂直な面であり、第１部分の第１出射面は凸球面、第２部分は第１部分の中心に位置、第２部分の第２入射面は光路に垂直な平面、第２部分の第２出射面の曲率は曲率よりも小さい第１部分の第１出射面の最大断面サイズ; 第２出射面の異なる断面寸法と照明効率および照明均一性との対応関係を計算することにより、第２出射面の最大断面サイズが得られます; 対応するサイズを選択します発光効率と発光均一度のいずれかが設定最小値に達したとき、第２出射面の最大断面サイズとして設定されます。	機械翻訳プラットフォームにおいて、冒頭の請求項番号が本文に組み込まれ「請求項1に記載の方法において」という不要な文言の湧き出しと派生させている。他のエンジンではそのような現象は見られない。

報告書　掲載項番		事例番号	採取文献	誤訳の内容	原文	校正文	機械翻訳プラットフォーム	エンジン①	エンジン②	エンジン③	エンジン④	短評
		3.7.1-2	CN107665903B ｾｸｼｮﾝH	請求項番号の誤解釈②	1.一种有机发光显示装置，包括：其特多个子像素的基板：分别设置在所述多个子像素中的多个第一电极：设置在每个所述第一电极的边缘上并与所述第一电极接触的第一电极保护图案：配置成与所述第一电极保护图案的一部分交叠以限定发光区域的堤部：位于所述第一电极上以对应于所述发光区域的有机发光层：以及位于所述有机发光层上的第二电极，其中，所述堤部包括：第一层，所述第一层配置成与每个所述第一电极的一部分交叠，使得相邻子像素的每个第一电极之间的区域被所述第一层填充：和第二层，所述第二层位于所述第一层上以对应于相邻子像素的每个第一电极之间的区域，以及其中，所述第一层是亲水绝缘材料并且所述第二层是疏水绝缘材料。	1.有機発光ディスプレイであって,以下を含む:複数のサブ画素を有する基板;それぞれ前記複数のサブ画素に設けられた複数の第一電極;前記第一電極の各々の縁部上に設置され,前記第一電極に接触する第一の電極パターン;前記第一電極保護パターンの一部と重なり発光領域を区画するように構成されるバンク;前記第一電極上に位置し前記発光領域に対応する有機発光層;および前記有機発光層上に位置する第二電極であって、ここで,前記バンクは以下を含む:第一層であって,前記第一層は各前記第一電極の一部と重なり合い,隣接するサブ画素の各第一電極間の領域が前記第一層によって充填されるように構成される;および第二層であって,前記第二層は前記第一層上に位置し隣接するサブ画素の各第一電極の間の領域に対応し,及びここで,前記第一層は親水性絶縁材料であり且つ前記第二層は疎水性絶縁材料である。	図1は,複数の副画素を有する基板と,前記複数のサブ画素のそれぞれに設けられた複数の第一の電極と,第一の電極の縁部上に配置され,前記第一の電極に接触する第一の電極パターンと,を含み,前記第一電極パターンの一部と重なることを特徴とする発光領域を区画するバンクに配置され;前記第一電極上には,前記発光領域の有機発光層であることを特徴とする有機発光表示装置。前記有機発光層の上に位置する第二電極にここで,前記現部は以下を含む:第一層であって,前記第一層は各前記第一電極の一部と重なり合うように配置され,隣接するサブ画素の各第一電極間の領域が前記第一層によって充填されるものこそして第二層であって,前記第二層は前記第一層の上に位置し隣接するサブ画素の各第一電極の間の領域に対応し,及びここで,前記第一層は親水性絶縁材料であり且つ前記第二層は疎水性絶縁材料である。	1. 有機発光表示装置であって、以下を含む:複数のサブ画素を有する基板;前記複数のサブ画素にそれぞれ設置される複数の第一電極;各前記第一電極の縁部に設置され且つ前記第一電極と接触する第一電極保護パターン;前記第1電極保護パターンの一部と重なるように配置されて発光領域を画定するバンク;前記第一電極に位置して前記発光領域に対応する有機発光層;及び前記有機発光層に位置する第二電極であって,そのうち,前記バンク部は以下を含む:第一層であって,前記第一層が各前記第一電極の一部と重なるように配置され,それにより隣接するサブ画素の各第一電極の間の領域が前記第一層で充填されるもの;第2の層は、隣接するサブピクセルの各第1の電極の間の領域に対応するように第1の層上に位置し、第1の層は親水性絶縁材料であり、第2の層は疎水性絶縁材料である。	1.有機発光表示装置であって,以下を含む:複数のサブ画素を有する基板;前記複数のサブ画素にそれぞれ設けられた複数の第一電極;各前記第一電極のエッジに設けられ且つ前記第一電極と接触する第一電極保護パターン;前記第一電極保護パターンの一部と重なるように配置されて発光領域を限定するバンク;前記第一電極に位置して前記発光領域に対応する有機発光層;及び前記有機発光層に位置する第二電極であって,ここで,前記バンクは以下を含む:第一層であって,前記第一層が各前記第一電極の一部と重なるように配置され,隣接するサブ画素の各第一電極の間の領域が前記第一層に充填されるもの;及び第二層であって,前記第二層が前記第一層に位置して隣接するサブ画素の各第一電極の間の領域に対応し,及びそのうち,前記第一層が親水性絶縁材料であり且つ前記第二層が疎水性絶縁材料であるもの。	1.有機発光表示装置であって,以下を含む:複数のサブ画素を有する基板;それぞれ前記複数のサブ画素に設置される複数の第一電極;各前記第一電極のエッジに設置され且つ前記第一電極と接触する第一電極保護パターン;前記第一電極保護パターンの一部と重なるように配置されて発光領域を限定するバンク部;前記第一電極に位置して前記発光領域に対応する有機発光層;及び前記有機発光層に位置する第二電極にここで,前記バンク部は以下を含む:第一層であって,前記第一層は各前記第一電極の一部と重なるように配置され,それにより隣接するサブ画素の各第一電極の間の領域は前記第一層により充填されるもの;及び第二層であって,前記第二層は前記第一層に位置して隣接するサブ画素の各第一電極の間の領域に対応し,且つここで,前記第一層は親水性絶縁材料であり且つ前記第二層は疎水性絶縁材料である。	複数のサブピクセルを有する基板と、複数のサブピクセルにそれぞれ配置された複数の第１の電極と、各第１の電極の端部に配置され、第１電極と接触する第１電極保護パターンと、発光領域を画定するために第１電極保護パターンの一部と重なるように構成されたバンクと、発光領域の有機発光に対応するように第１電極上に位置するバンクと、前記バンクは、サブピクセルの各第１電極間の隣接領域が充填されるように、前記第１電極の各々の一部と重なるように構成された第１層を含む：第１の層と、隣接するサブピクセルの各第１の電極に対応する第１の層上の第２の層とを含み、第１の層は親水性の絶縁材料であり、第２の層は疎水性の絶縁材料である。	上例と同様、機械翻訳プラットフォームでは請求項番号が本文に組み込まれ「図1は」と誤訳されたが、他のエンジンでは発生しなかった。ただしエンジン④は主題の「有機発光ディスプレイ」が訳抜けしており、また請求項番号も出力されていない。
	3.7.2. 発明の主題を文末に出力する傾向	3.7.2-1	CN107920803B ｾｸｼｮﾝA	主題を文末へ再出力①	1.一种超声波阵列振荡器，包括：基板：	1.超音波アレイ振動子であって,以下を含む:基板;	1.超音波アレイ発振器であって,以下を含む:基板;	1. 超音波アレイ発振器であって、以下を含む:基板;	1.超音波アレイ発振器であって,以下を含む:基板;	1.超音波アレイ発振器であって,以下を含む:基板;	超音波アレイ発振器であって、基板と、	機械翻訳プラットフォームにおいて、主題を文末に配置する日本の請求項記載様式を学習した影響と、長大な請求項を複数に分割して翻訳する処理、先行する行の情報を後続行に引き継げない制約が相まって、一構成要素にすぎない4行目の「半導体チップ」が、あたかも請求項全体の主題のように文末に再配置された事例。他エンジンにも同様の現象が見られ、エンジン①と③はともに「半導体チップと」という中途半端な形で再配置されている。エンジン②と④は、原文・校正文とは構成が異なるが、主題の文末配置は行われていない。
					彼此独立的多个超声波振荡器模块，所述多个超声波振荡器模块构成阵列：	互いに独立した複数の超音波振動子モジュールであって,前記複数の超音波振動子モジュールがアレイを構成するもの;	互いに独立した複数の超音波振動子モジュールであって,前記複数の超音波振動子モジュールがアレイを構成するもの;	アレイを構成する互いに独立した複数の超音波発振器モジュール;	互いに独立した複数の超音波発振器モジュール,前記複数の超音波発振器モジュールは、アレイを構成する;	互いに独立した複数の超音波発振器モジュールであって、アレイを形成する、複数の超音波発振器モジュールと	互いに独立した複数の超音波発振モジュールであって、アレイを形成する複数の超音波発振モジュールと、	
					其中所述多个超声波振荡器模块的每一个包含：超声波振荡器：	ここで前記複数の超音波振動子モジュールの各々は以下を含む:超音波振動子;	ここで前記複数の超音波振動子モジュールの各々は以下を含む:超音波振動子;	そのうち前記複数の超音波発振器モジュールのそれぞれは以下を含む:超音波発振器;	前記複数の超音波発振器モジュールの各々は、超音波発振器を含み、;	前記複数の超音波発振器モジュールの各々は、超音波発振器を含み	複数の超音波振動子モジュールのそれぞれは、超音波振動子を含む。	
					和半导体芯片，所述半导体芯片形成用于所述超声波振荡器的阻抗匹配电路，所述半导体芯片接合到所述超声波振荡器并设置在所述基板和所述超声波振荡器之间，并且所述半导体芯片将设置在所述基板上的布线电连接至所述超声波振荡器。	および超音波振動子用インピーダンス整合回路を形成する半導体チップであって,前記半導体チップは前記超音波振動子に接合され,且つ前記基板と前記超音波振動子との間に配置され,且つ前記半導体チップは,前記基板上に配置された配線を前記超音波振動子に電気的に接続するもの。	そして超音波振動子用インピーダンス整合回路を形成する半導体チップであって,前記半導体チップは前記超音波振動子に接合されて前記基板と前記超音波振動子との間に配置され,前記半導体チップは,前記基板上に配置された配線を前記超音波振動子に電気的に接続することを特徴とする半導体チップ。	前記超音波発振器用のインピーダンス整合回路を形成する半導体チップであって、前記超音波発振器に接合され、前記基板と前記超音波発振器との間に設けられ、前記基板に設けられた配線と前記超音波発振器とを電気的に接続する半導体チップと。	半導体チップであって、前記半導体チップは、前記超音波発振器に用いられるインピーダンス整合回路を形成し、前記半導体チップは、前記超音波発振器に接合された前記基板と前記超音波発振器との間に設けられ、前記半導体チップは、前記基板上に設けられた配線を前記超音波発振器に電気的に接続する。	前記超音波振動子用のインピーダンス整合回路を形成する半導体チップであって、前記超音波振動子に接合され、前記基板と前記超音波振動子との間に配置され、前記基板に設けられた配線と前記超音波振動子とを電気的に接続する半導体チップと	超音波振動子のインピーダンス整合回路を形成する半導体チップと、超音波振動子に接合された基板と超音波振動子との間に設けられ、半導体チップは、基板に設けられた配線と超音波振動子とを電気的に接続する。	
		3.7.2-2	CN106394270B ｾｸｼｮﾝB	主題を文末へ再出力②	1.一种车辆无线充电系统，包括：	1.車両無線充電システムであって,以下を含む:	1.車両無線充電システムであって,以下を含む:	1. 車両無線充電システムであって、以下を含む:	1.車両無線充電システムであって,以下を含む:	1.車両の無線充電システムであって,以下を含む:	1. 車両ワイヤレス充電システムであって、	機械翻訳プラットフォームの翻訳文は、上例と同様、4行目に記載された「車両電力変換装置」が主題であるかのように文末に再配置されている。また、全体的な文意もきわめて不正確である。他の各エンジンでは主題の再配置は発生しなかったが、エンジン①、②、④は機械翻訳プラットフォームと同レベルの低品質な翻訳文となっており、かなり難易度の高い文であるようだ。これらに比べ、エンジン③は、語順を大幅に組み替えているものの、概して正確な翻訳である。
					逆变器：	インバータ;	インバータ;	インバータ;	インバータ;	インバータと	インバーター;	
					至少一个控制器，被配置为：	少なくとも一つのコントローラであって,以下のように構成される:	少なくとも一つのコントローラであって,以下のように構成されるもの:	少なくとも1つのコントローラであって,以下のように構成されるもの;	少なくとも1つのコントローラであって,以下のように構成されるもの;	少なくとも1つのコントローラであって	次のように構成された少なくとも１つのコントローラ:	
					根据到车辆电力转换器的电压输入与电压基准之间的误差改变所述逆变器的输出电压的频率，以控制所述电压输入，从而使得所述逆变器的输出处的阻抗相位角趋向于预定角度并达到车辆电力转换器的输出处的功率需求，并且基于加权的目标函数的变化率改变所述逆变器的输出电压的频率，其中，所述电压基准是基于所述功率需求、车辆电力转换器的功率输出、所述阻抗相位角以及所述预定角度的，所述加权的目标函数被配置为减小车辆电力转换器的输出功率误差和所述逆变器的输出处的阻抗相位角误差。	車両電力変換器への電圧入力と電圧基準との誤差に基づき前記インバータの出力電圧の周波数を変えて,前記電圧の入力を制御し,それにより前記インバータの出力におけるインピーダンスの位相角が所定の角度に近づき,かつ車両電力変換器の出力における電力需要に達し,かつ加重目的関数の変化率に基づき前記インバータの出力電圧の周波数を変更し,ここで前記電圧基準は前記電力要件,車両の電力変換器の電力出力,前記インピーダンス位相角および所定の角度に基づくもので,前記加重目的関数は車両の電力変更器の出力電力の誤差および前記インバータの出力におけるインピーダンス位相角の誤差を低減するように構成される。	前記インバータの出力電圧の周波数は,前記インバータの出力におけるインピーダンス位相角が所定角度になるように前記インバータの出力電圧の周波数を変化させて車両電力コンバータ出力の電力需要に到達するように車両電力変換装置の電圧入力と電圧基準との誤差に応じて変更され,前記電力要求,自動車電力変換装置の電力出力,前記インピーダンス位相角及び前記所定角度に基づいて,車両電力コンバータ出力電力誤差及び前記インバータ出力のインピーダンス位相角誤差を低減するように構成されることを特徴とする車両電力変換装置。	インバータの出力電圧の周波数は、車両電力変換器への電圧入力と、電力需要、車両電力変換器の電力出力、インピーダンス位相角、および所定の角度に基づく加重目的関数の変化率に基づいて、車両電力変換器の出力電力誤差およびインバータの出力におけるインピーダンス位相角誤差を低減するように構成された電圧基準との間の誤差に基づいて、インバータの出力におけるインピーダンス位相角が所定の角度に近づき、車両電力変換器の出力における電力需要に達するように、電圧入力を制御するように変更される。	車両電力変換器への電圧入力と電圧基準との間の誤差に応じて、インバータの出力におけるインピーダンス位相角が、電力需要、車両電力変換器の電力出力、インピーダンス位相角、および車両電力変換器の出力電力誤差およびインバータの出力におけるインピーダンス位相角誤差を低減するように構成される、車両電力変換器の出力における電力需要に到達し、重み付けされた目的関数の変化率に基づいて、インバータの出力電圧の周波数を変化させるように、インバータの出力電圧の周波数が変更される。	車両電力変換器への電圧入力と、電力需要、車両電力変換器の電力出力、インピーダンス位相角、および所定の角度に基づく電圧基準との間の誤差に応じて、インバータの出力電圧の周波数を変化させて、インバータの出力におけるインピーダンス位相角が所定の角度に近づき、車両電力変換器の出力における電力需要に達するように電圧入力を制御し、車両電力変換器の出力電力誤差およびインバータの出力におけるインピーダンス位相角誤差を低減するように構成された重み付き目的関数の変化率に基づいて、インバータの出力電圧の周波数を変化させる。	インバータの出力電圧の周波数は、車両パワーコンバータへの電圧入力と電圧入力を制御するための電圧基準との間の誤差に従って変化し、インバータの出力におけるインピーダンス位相角が、重み付けされた目的関数の変化率に基づくインバータ出力電圧の周波数。ここで、電圧基準は電力需要に基づいています。車両電力コンバータの電力出力の重み付けされた目的関数、インピーダンス位相角、角度は、車両電力コンバータの出力電力誤差と、インバータの出力におけるインピーダンス位相角誤差を低減するように構成されています。	
	3.7.3. 発明構成要素の訳ゆれ	3.7.3-1	CN107074359B ｾｸｼｮﾝB	発明構成要素「推進器」の訳ゆれ	2.根据权利要求1所述的航空器，其中所述止挡构件被配置成使所述推进器通过以下动作中的至少一者停止：从所述马达移除电流；使至所述马达的所述电流的极性逆转；将止挡杆部署至所述马达的转子中；或使所述推进器从所述马达脱离。	2.請求項1に記載の航空機であって,ここで前記ストッパ部材は前記プロペラが以下の動作のうちの少なくとも一つによって停止するように配置される:前記モータから電流を除去する;前記モータへの前記電流の極性を反転させる;ストッパロッドを前記モータの回転子に配置する;又は前記プロペラを前記モータから離脱させる。	2.請求項1に記載の航空機であって,ここで前記ストッパ部材は前記スラスタが以下の動作のうちの少なくとも1つによって停止するように配置される:前記モータから電流を除去する;前記モータへの前記電流の極性を反転させる;ストッパロッドを前記モータの回転子に配置する;又は前記推進器を前記モータから離脱させる。	2. 請求項1に記載の航空機であって、前記ストッパ部材は、前記ブッシャーである：前記モータから電流を除去することと;ストッパロッドを前記モータの回転子内に配置する;又は前記プロペラを前記モータから離脱させる。	2.請求項1に記載の航空機であって,そのうち前記ストッパ部材は前記推進器が以下の動作のうちの少なくとも一つによって停止するように配置される:前記モータから電流を除去する;前記モータまでの前記電流の極性を逆転させる;ストッパロッドを前記モータの回転子に配置する;又は前記推進器を前記モータから離脱させる。	前記停止部材は、前記モータから電流を除去すること、前記モータへの前記電流の極性を逆転させること、前記モータのロータ内に停止バーを展開すること、または前記モータから前記プロペラを係合解除することのうちの少なくとも1つによって、前記プロペラを停止するように構成される、請求項1に記載の航空機。	前記停止部材は、前記モータから電流を除去すること、前記電流の極性を逆にすること、前記モータのロータ内に停止ロッドを配備すること、または係合を解除することのうちの少なくとも１つによって、前記プロペラを停止するように構成される、請求項１に記載の航空機。モーターからプロペラ。	機械翻訳プラットフォームでは、「推進器」が「スラスタ」と「推進器」に訳ゆれしている。エンジン①も「ブッシャー」と「プロペラ」で訳ゆれが発生。一方、エンジン②は「推進器」、エンジン③と④は「プロペラ」で統一されている。ただし、本案件の全請求項で調査したところ、「推進器」は37回出現し、機械翻訳プラットフォームではスラスタ×14、推進器×10、ブッシャ×6、プロペラ×4、推進装置×2、インペラ×1であったのに対し、エンジン①ではプロペラ×19、推進器×9、ブッシャ×7、スラスタ×2、エンジン③ではプロペラ×26、スラスタ×6、インペラ×4、エンジン②と④も同じような訳ゆれが生じており、案件単位では各エンジンとも訳ゆれが生じていた。
	3.7.5. 事例集	3.7.5-1	CN106659153B ｾｸｼｮﾝA	文末への不要な文言の出力	e)相对于总的组合物为50g/kg至250g/kg的固体缓冲剂，其中所述固体缓冲剂为硫酸铵或柠檬酸；和	e)全組成物に対して50g/kg～250g/kgの固体緩衝剤であって,ここで,固体緩衝剤は硫酸アンモニウムまたはクエン酸である;および	e)全組成物に対して50g/kgから250g/kgまでの固体緩衝剤(ここで、固体緩衝剤は硫酸アンモニウムまたはクエン酸である);および固体緩衝剤。	e) 全組成物に対して50g/kg～250g/kgの固体緩衝液であって,ここで前記固体緩衝液は硫酸アンモニウム又はクエン酸である;そして	e)固体緩衝剤が硫酸アンモニウムまたはクエン酸である、全組成物に対して50g/kg~250g/kgの固体緩衝剤;および	e)全組成物に対して50g/kg～250g/kgの固体緩衝剤であって、硫酸アンモニウムまたはクエン酸である固体緩衝剤;	ｅ）全組成物に対して５０ｇ／ｋｇ～２５０ｇ／ｋｇの固体緩衝剤であって、固体緩衝剤が硫酸アンモニウムまたはクエン酸である：および	機械翻訳プラットフォームにおいて、主題を文末に再配置する傾向が不適切に働いて不備を生じた事例。他のエンジンでは同様の不具合は生じなかった。
					f)崩解剂，其中所述崩解剂为聚醛共聚物。	f)崩壊剤であって,ここで前記崩壊剤は尿素ホルムアルデヒドコポリマーである。	f)崩壊剤が尿素尿素コポリマーである、崩壊剤。	f) 崩壊剤であって,ここで前記崩壊剤は尿素アルデヒド共重合体である。	f)崩壊剤であって,ここで前記崩壊剤は尿素ホルムアルデヒド共重合体である。	f)尿素-ホルムアルデヒドコポリマーである崩壊剤	ｆ）崩壊剤であって、崩壊剤が尿素－ホルムアルデヒドコポリマーである、崩壊剤。	