

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

中国審決の機械翻訳精度の調査・分析

最終報告書

令和4年3月22日

一般財団法人 日本特許情報機構

内容

報告書概略.....	1
I. 事業実施報告.....	17
1. 本事業の概要.....	19
1.1. 本事業の背景・目的.....	19
1.2. 本事業の内容.....	20
1.3. 本事業の成果物.....	20
1.3.1. 中国審決翻訳データ.....	20
1.3.2. 中国審決コーパスデータ.....	21
1.3.3. 中国審決対訳辞書データ.....	22
1.3.4. 中国審決汎用文例リスト.....	23
1.3.5. 中国審決翻訳データベース.....	23
II. 調査報告.....	25
2. 調査の概要.....	27
2.1. 調査の方法.....	27
2.2. 使用した機械翻訳システム.....	28
2.2.1. 汎用 NT と特許 NT.....	28
2.2.2. 「みんなの自動翻訳」のバージョンアップについて.....	28
2.2.3. 「みんなの自動翻訳」の大規模アップデート.....	29
2.3. 自動評価と人手評価のタイミング.....	29
2.4. 評価用コーパス.....	30
2.4.1. 案件単位の選定.....	30
2.4.2. 復審請求／無効宣告請求について.....	31
2.4.3. 法区分について.....	31
2.4.4. 技術分野について.....	31
2.4.5. 適用条文について.....	31
2.4.6. 一部案件における「特許請求の範囲」部分の除外.....	32
2.4.7. 選定結果.....	32
2.4.8. 人手評価対象文について.....	32
3. 自動評価結果とその分析.....	33
3.1. 自動評価の概要.....	33
3.2. BLEU と RIBES.....	33
3.3. 各回の自動評価スコアの集計結果.....	35
3.3.1. 評価用コーパス全件の平均スコア.....	35
3.3.2. 文単位のスコア向上／低下カウント結果.....	40

3.3.3. 審決種別、適用条文別、当事者別の集計結果	43
3.3.3.1. 審決種別ごとの集計	43
3.3.3.2. 適用条文による集計	45
3.3.3.3. 技術分野間の差異	46
3.4. 実例を用いた機械翻訳結果の定点観測	48
3.4.1. 実例 3-1：文番号#1、特許 NT	48
3.4.2. 実例 3-2：文番号#1110、汎用 NT／特許 NT	61
3.4.3. 実例 3-3：文番号#683、汎用 NT	66
3.4.4. 実例 3-4：文番号#1995、汎用 NT	74
3.4.5. 実例 3-5：文番号#1155、汎用 NT	83
3.4.6. 実例 3-6：文番号#547、汎用 NT [翻訳不能文とその改善]	93
4. 人手評価結果とその分析	99
4.1. 人手評価の概要	99
4.2. 人手評価観点	99
4.3. 人手評価者	99
4.4. 人手評価の対象文	100
4.5. 各評価項目の集計結果	101
4.5.1. 内容伝達レベルの評価	101
4.5.2. 個々の翻訳エンジンの学習前後の相対評価	105
4.5.3. 両エンジン間の機械翻訳文の相対評価	108
4.5.4. 実例を用いた機械翻訳文の変遷の分析	109
4.5.4.1. 実例 4-1：文番号#1494、汎用 NT の分析	109
4.5.4.2. 実例 4-2：文番号#704、汎用 NT の分析	114
4.5.4.3. 実例 4-3：文番号 # 830、汎用 NT の分析	119
4.5.4.4. 実例 4-4：最終学習後に内容伝達レベルが最も低かった実例の分析	123
4.5.5. 「エラーのカテゴリ別カウント」結果	129
4.5.5.1. 「特許誤訳」のカウント結果	129
4.5.5.2. 「技術／一般誤訳」のカウント結果	135
4.5.5.3. 「未知語」のカウント結果	141
4.5.5.4. 「係り受けエラー」のカウント結果	144
4.5.5.5. 「日本語表現エラー」のカウント結果	147
4.5.5.6. 「欠落エラー」のカウント結果	149
4.5.5.7. 「湧き出しエラー」のカウント結果	154
4.5.5.8. 「数値エラー」のカウント結果	161
4.5.5.9. 「記号エラー」のカウント結果	168
4.5.5.10. 「その他エラー」	168

5. 中国審決への機械翻訳の適用に関する分析・考察	173
5.1. 全体的な品質改善効果について	173
5.2. 特許・審決用語の訳質改善について	174
5.3. 技術用語・一般用語の網羅性について	175
5.4. 技術用語・一般用語の不安定な変化について	176
5.5. 「専利」の訳し分けに関する課題	181
5.6. 長文における翻訳文の途絶について	183
5.7. 汎用 NT における日付エラーについて	183
5.8. 汎用 NT と特許 NT の優劣	184
5.9. 自動評価指標における第 2 回学習以降のスコアの停滞について	185
5.10. 本事業で達成された中国審決の機械翻訳品質について	186

報告書概略

本事業の背景・目的

中国における特許・実用新案の出願件数の急増に伴い、中国での特許等の審判事件の処理件数も増加しており、我が国や欧米の処理件数を上回ってなお増加傾向にある。今後は我が国企業が関連する中国での審判事件が一層増加することが想定され、審決情報の重要性が高まっている。しかしながら、中国審決情報は一次情報としては中国語でのみ提供されており、我が国企業からは当該中国審決情報の日本語翻訳文の提供が強く要望されている。よって、我が国企業の円滑な権利取得と競争力向上の観点から、我が国企業による中国審決の分析や中国での審判事件への対応を効率化するため、中国審決情報の日本語への機械翻訳精度を向上させる必要がある。

現在、最も注目されている機械翻訳方式は、ニューラル機械翻訳である。ニューラル機械翻訳は、ニューラルネットワークによる深層学習技術を利用し、原文とその翻訳文とからなる対訳文のデータ（対訳コーパス。本報告書では単に「コーパス」ともいう。）を機械学習することにより翻訳を行うものである。従前の機械翻訳よりも高精度に自然な翻訳文を生成することができるが、専門性の高い文書を精度よく翻訳させるには、その文書に特有の用語や文体の学習が必要である。中国審決には専門的な審決用語や表現が多く含まれるため、高精度な機械翻訳文を提供するためには、実際の中国審決文に由来する対訳コーパスを十分な量作成し、翻訳エンジンに学習させる必要がある。

本事業では、中国審決を人手により翻訳して和訳文を提供するとともに、将来的な中国審決の機械翻訳文の提供に向け、和訳文から中国審決由来の対訳コーパスを作成する。また、ある用語を常に特定の訳語に翻訳できるようにするとともに、対訳コーパスを利用しない従来方式の機械翻訳へも対応すべく、対訳辞書も併せて作成する。これにより、我が国ユーザの中国審決に対するアクセス性を改善し、中国における権利取得の業務を効率化させることを目指す。

事業内容

本事業では、中国審決 19,182 件の高精度な全文人手翻訳データである「中国審決翻訳データ」、この翻訳データと中国審決原文とを文単位の対訳データ形式に編集した 2,000,059 文対（重複しない文対単位。重複文対を含めると 2,694,613 文対）の対訳コーパス「中国審決コーパスデータ」、同コーパスから抽出した 100,000 語の「中国審決対訳辞書データ」、中国審決の書誌事項と中国審決翻訳データ中の「決定の要旨」の翻訳文から生成した「中国審決翻訳データベース」19,182 件、中国審決コーパスデータ上で出現頻度 10 回以上の文対のみを抽出した「中国審決汎用文例リスト」（5,412 文対を収録）の 5 種類の言語資源を作成した。

本事業ではさらに、作成した中国審決コーパスデータをニューラル機械翻訳「みんなの自動翻訳」の2つの翻訳エンジン（汎用 NT／特許 NT）に段階的に学習させ¹、共通の「評価用コーパス（1,994 文）」を入力文に用いて、得られた都度の機械翻訳結果の品質を評価することで、学習の進捗に伴う中国審決の機械翻訳精度の改善の度合いや変化の傾向、実用性等を調査した。

事業実施結果

本事業で作成した主な言語資源は以下のとおりである。

- ・中国審決翻訳データ 19,182 文献
- ・中国審決コーパスデータ 2,000,059 文対（重複しない文対単位）
2,694,613 文対（重複文対含む）
- ・中国審決対訳辞書データ 100,000 語
- ・中国審決翻訳データベース 19,182 件
- ・中国審決汎用文例リスト 5,412 文対

また、本事業の一環として実施した「みんなの自動翻訳」の機械翻訳品質の評価結果及びその分析・考察内容について、最終報告書にまとめた。次章より、その主な内容を記す。

¹ 中国審決コーパスデータは「重複しない文対」単位で約 10 万件ずつ、全 20 回にわたり段階的に学習に供した。ただし、学習にはデータ中における出現頻度も重要であるとの考えから、学習データ自体は重複文対を含めた全件を使用した。このため各回の学習データの実件数は約 14 万件程度となる。

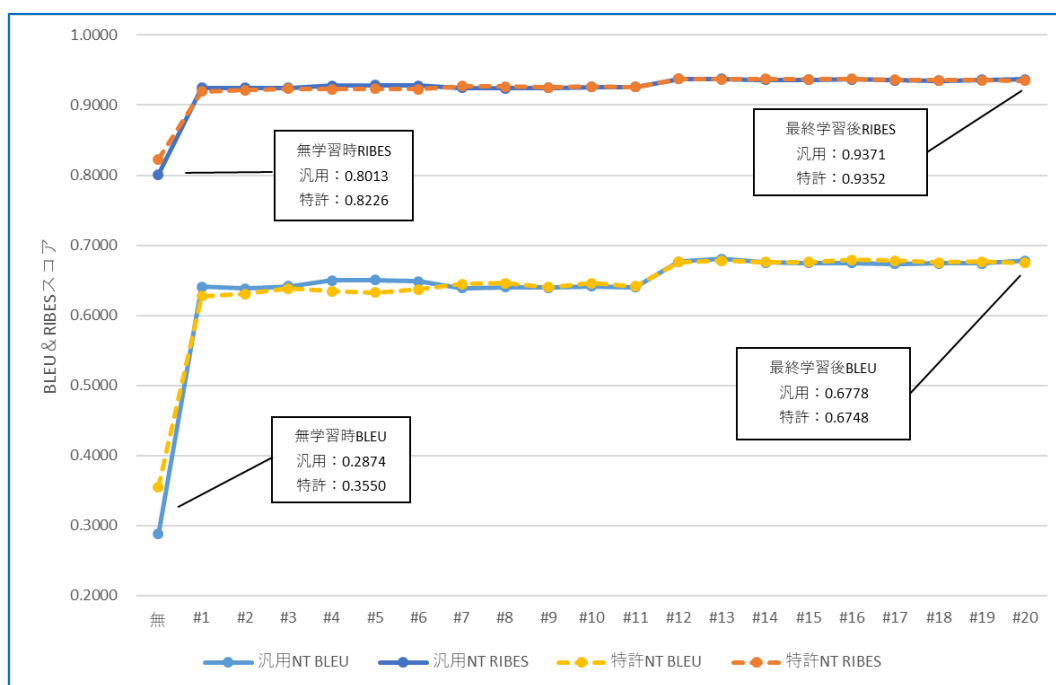
機械翻訳品質の評価結果

「みんなの自動翻訳」の機械翻訳品質の評価には、自動評価と人手評価を併用した。以下、それぞれの評価結果を示す。

[自動評価結果]

自動評価は、無学習時及び各回学習後（＝学習データを約 10 万文対適用する都度）の全 21 回実施した。評価指標には BLEU と RIBES を併用し、各学習段階における評価用コーパス各文の評価スコアを取得した。以下、評価用コーパス全件の平均スコアの推移を示す。

評価用コーパス全件の BLEU、RIBES 平均スコア推移



上記グラフ中、主な学習回の BLEU スコア（「B」と表記する。）、RIBES スコア（「R」と表記する。）を下表に示す。

評価用コーパス全件の BLEU、RIBES 平均スコア一覧（主要学習回を抜粋）

学習回		無学習	第 1 回	第 5 回	第 10 回	第 12 回	第 15 回	第 20 回
汎用 NT	B	0.2874	0.6406	0.6507	0.6417	0.6766	0.6747	0.6778
	R	0.8013	0.9247	0.9283	0.9260	0.9374	0.9363	0.9371
特許 NT	B	0.3550	0.6273	0.6327	0.6456	0.6762	0.6761	0.6748
	R	0.8226	0.9198	0.9283	0.9264	0.9372	0.9367	0.9352

中国審決コーパスデータ全件学習後の全体平均スコアは、汎用 NT は BLEU が 0.6778 (無学習時から +0.3904)、RIBES が 0.9371 (同+0.1358)、特許 NT は BLEU が 0.6748 (同+0.3198)、RIBES が 0.9352 (同+0.1126) といずれも向上した。なお、グラフに示したとおり、特に第 1 回学習後に大幅なスコア向上が果たされた一方、以後の各回はおおむね微増と微減を繰り返す形となった。唯一、第 12 回学習後のみは両エンジン、両指標とも明瞭なスコア向上が見られた。これは当時実施された「みんなの自動翻訳」の大規模アップデートの効果と見られ、以降も第 12 回時に上昇したスコア水準が保たれた。

なお、本調査では、こうした自動評価結果の妥当性の確認や変動要因の分析のため、サンプリングした実例による各回機械翻訳結果の変化状況の定点観測をあわせて実施した。そこで判明した事項については後述する (⇒[評価結果の分析及び考察](#))

[人手評価結果]

人手評価は、評価用コーパスから選定した 100 文のみを対象とし、無学習時及び学習データを約 50 万文適用する都度 (学習 5 回ごと) の全 5 回実施した。以下、評価観点ごとの評価結果を示す。

①内容伝達レベルの評価

機械翻訳文から原文の内容が正確に把握できるかを以下の 5 段階のレベルで評価し、各文に付されたレベル値を集計した。

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

下表に評価対象 100 文の各回における「内容伝達レベル」の平均値を示す。

「内容伝達レベル」評価スコアの集計結果

エンジン	無学習	第 5 回	第 10 回	第 15 回	第 20 回	改善度
汎用 NT	2.41	4.42	4.32	4.61	4.65	+2.24
特許 NT	2.18	4.28	4.30	4.66	4.61	+2.43

汎用 NT では無学習時の 2.41 から第 20 回学習後には 4.65 へ (+2.24)、特許 NT では 2.18 から 4.61 へ (+2.43) とそれぞれ大きく向上した。両エンジンとも学習初期（第 5 回時）に最も大幅な改善を見せ（汎用 NT:+2.01、特許 NT:+2.10）、「みんなの自動翻訳」の大規模アップデートを経た第 15 回学習後にもそれに次ぐ改善があり（汎用 NT:+0.29、特許 NT:+0.36）、最終学習となる第 20 回学習後もその水準が維持された。こうした推移は自動評価スコアと一致している。

② 2つの翻訳エンジンの相対評価

汎用 NT と特許 NT それぞれの各回学習後の機械翻訳結果を文ごとに比較し、その優劣を判定した。以下、各回の集計結果を示す。

各回学習後の両エンジン間の翻訳品質の「相対評価」結果

	第 5 回	第 10 回	第 15 回	第 20 回
汎用 NT が優れる	43	28	8	12
両エンジン同等	37	50	79	79
特許 NT が優れる	20	22	13	9

上表のとおり、学習させる中国審決コーパスデータを増量するにつれて機械翻訳品質が「両エンジン同等」に収れんしていき、最終学習後には個々の文単位でも全体の約 8 割に相当する 79 文が品質上の優劣なしと判定された。残り約 2 割の優劣も汎用 NT が 12 文、特許 NT が 14 文と拮抗しており、コーパス学習後の両エンジンの翻訳品質は同程度と見なせる。

③ エラーのカテゴリ別カウント

機械翻訳結果に頻出する典型的なエラーを 10 種類にカテゴリライズし、該当するエラーが発生しているか否かを文単位でカウントした。以下、カテゴリ別の集計結果を示す。

無学習及び各回学習後の「エラーのカテゴリ別カウント」集計結果

カテゴリ	エンジン	無学習	第 5 回	第 10 回	第 15 回	第 20 回	エラー解消
特許誤訳	汎用	91	5	5	15	9	-82
	特許	90	21	13	16	16	-74
技術／一般誤訳	汎用	40	21	18	11	13	-27
	特許	37	20	16	11	11	-26
未知語	汎用	19	1	1	1	0	-19
	特許	2	1	1	1	0	-2
係り受けエラー	汎用	42	19	17	7	5	-37

	特許	49	15	20	4	6	-43
日本語表現エラー	汎用	43	24	23	13	10	-33
	特許	42	18	21	17	14	-28
欠落	汎用	38	10	15	2	4	-34
	特許	45	13	13	1	2	-43
湧き出し	汎用	14	3	6	0	0	-14
	特許	30	9	8	0	0	-30
数値エラー	汎用	23	11	10	5	5	-18
	特許	22	12	10	5	5	-17
記号エラー	汎用	8	2	1	0	0	-8
	特許	8	1	1	0	0	-8
その他エラー	汎用	2	0	0	0	0	-2
	特許	4	2	1	0	0	-4

各カテゴリとも無学習時に比べてエラーの発生文数が減少していることが示された。なかでも顕著な改善がなされたのが特許・審決用語が誤訳されている文をカウントした「特許誤訳」のカテゴリであり、無学習の時点では評価対象 100 文の大多数において何らかの特許誤訳が発生していたところ、中国審決コーパスデータの学習により大幅に改善したことが示された。

評価結果の分析及び考察

得られた評価結果及び実際の機械翻訳結果を分析し、「みんなの自動翻訳」に「中国審決コーパスデータ」を適用することによる品質改善効果について考察した結果を本項に記す。

[全体的な品質改善効果について]

本事業は、中国審決は専門性が高く、デフォルト状態のニューラル機械翻訳では十分な精度の翻訳結果は得られないとの前提のもと、中国審決由来の対訳コーパスを大量に作成し、これを機械翻訳エンジンに学習させることで品質改善を図ることを目的に実施された。

この観点から各評価結果を考察すると、まず評価用コーパス全件を対象とした自動評価においては、BLEU、RIBES の両指標とも、中国審決コーパスデータを学習させた結果、無学習時点に比べて大幅にスコアが向上しており、コーパスの学習による翻訳精度改善を示す結果となった。そして、評価用コーパスのうち 100 文を対象とした「内容伝達レベル」の人手評価においても同様の結果が示され、自動評価結果の妥当性が裏付けられた。

具体的な翻訳品質改善の度合いは、「内容伝達レベル」の平均スコアに端的に示される。

汎用 NT では無学習時のスコア 2.41 から第 20 回学習後には 4.65 へ、特許 NT では 2.18 から 4.61 へとそれぞれ向上したが、これは前掲の「内容伝達レベル」のスコア定義にあてはめると、無学習時には「いくつかの重要情報は正確に伝達されている（20%～）」程度であった機械翻訳品質が、最終学習後には「ほとんどの重要情報は正確に伝達されている（80%～）」以上のレベルにまで精度向上したことを意味している。

この結果から、上述した本事業の前提である、中国審決はデフォルト状態のニューラル機械翻訳では十分な精度の翻訳結果は得られないこと、及び中国審決由来の対訳コーパスを大量に学習させることで品質改善が果たされることの双方が実証された。より具体的には、本事業で作成した「中国審決コーパスデータ」全件を「みんなの自動翻訳」に学習させることにより、中国審決の機械翻訳品質を、重要情報の 20～40%程度しか伝達されないレベルから、重要情報の 80%以上が正確に伝達されるレベルまで改善できることが証明された²。

〔特許・審決用語の訳質改善について〕

中国審決コーパスデータを学習させる主目的は、機械翻訳エンジンを中国審決特有の用語や表現に対応させることにある。この目的がどの程度達成されたかは、人手評価における「エラーのカテゴリ別カウント」の「特許誤訳」において把握できる。

「特許誤訳」では特許・審決用語及び表現が誤訳されている文をカウントしている。無学習の時点では全 100 文中、汎用 NT が 91 文、特許 NT が 90 文と、大半の文で何らかの特許誤訳が発生していたのに対し、第 20 回学習後はそれぞれ 9 文、16 文と大幅な改善を見せた³。

サンプリングによる各学習回での機械翻訳結果の定点観測でも、大半の特許・審決用語は初回学習時から中国審決コーパスデータで使用されている適訳語に改善され、以降の各学習回後でも常時その訳語が採用され続けていた。こうした特許・審決用語の訳質改善が、前項に示した全体的な品質改善効果の最大の要因であり、本事業で作成した中国審決コーパスデータがその目的を十分に果たしていることが確認できた。

² なお、無学習時と第 20 回学習後とは「みんなの自動翻訳」自体のバージョンが異なり、品質改善にはこうしたエンジン自体の性能向上が寄与した部分も含まれている。

³ 学習後期まで残存した特許誤訳の大部分は「専利」の訳し分けの失敗であった。この課題については後述する

[技術用語／一般用語の不安定な変化について]

審決書類は発明や先行技術の技術内容に深く言及するため、技術／一般用語も多数含む。このため、技術用語や一般用語の翻訳精度も機械翻訳品質に影響する。中国審決には請求項など原出願の摘記箇所も多く含むため、中国審決コーパスデータにも多くの技術用語が含まれる。とはいえ、中国審決コーパスデータは主に特許・審決用語の学習のためのものであり、あらゆる技術／一般用語を網羅してはいない。このため技術／一般用語は、基本的には機械翻訳システム自身の基礎学習によるデフォルトの翻訳精度に大きく依存する。その点、「みんなの自動翻訳」の技術／一般用語の翻訳品質は、例えば内容伝達レベルの評価結果が最終的には「ほとんどの重要情報は正確に伝達されている（80%～）」水準に達していることや、エラーカテゴリ「技術／一般誤訳」の最終的なカウント数が全体の10%強に抑えられていることから判断して、一定の水準にあると考えてよいであろう。

ただし、事例の定点観測においては、学習初期から中国審決コーパスデータ由来の訳語で常時安定する傾向であった特許・審決用語と異なり、技術用語や一般用語の中には、出現の都度、あたかも複数の訳語候補からランダムに選定されているかのような予測不可能な変化を見せるものが散見された。例を挙げれば、評価用コーパス中の一文（事例#1995）においては、「匹配的（見合う）」という一語に対し、汎用 NT では初回学習後から第10回学習後にかけて「一致する（第1、2、5、7回）」、「マッチングする（第3、4、8、10回）」、「適合する（第6、9回）」という3種の訳語が入れ替わり採用されている。この間、特許 NT も第1回の「マッチする」を除き「一致する（第2、4、5、7回）」「マッチングする（第6、9回）」「適合する（第3、8、10回）」と同じ3種の訳語が採用されていたが、2エンジンの訳語が同一回で一致したのは第2回学習後と第5回学習後と第7回学習後の3回のみであった。

こうした訳語のランダム的な変化は、調査期間中に実施された「みんなの自動翻訳」の大規模アップデートを機にやや抑制されたものの、最終学習時まで残存した。このため、現状では同一文献中での技術／一般用語の訳ゆれが生じる懸念がある。事実、用語によっては同一文献中での訳ゆれも検出されており、実用時に読み手の混乱を招くおそれがある。

学習初期から適訳語で常時安定した特許・審決用語と、全期間にわたり不安定に変化した技術／一般用語との傾向の違いは、特許・審決用語の多くが中国審決コーパスデータで使用されている単一の訳語候補しか持たないのに対し、技術用語や一般用語には複数の訳語を持つものも多く、複数の訳語候補が「みんなの自動翻訳」の学習データ（あらかじめエンジンが学習していたものを含む）中に存在しているためと考えられる。

ニューラル機械翻訳では、ある用語について学習データ中に複数の訳語が存在する場合、

それぞれの訳語に対し、その訳語が使われた各文例における周辺情報に基づいた多次元ベクトルを付与する形で学習を行う。そして、実際の翻訳で訳語を選定する際は、入力文におけるその用語の周辺情報と、学習データ中の各訳語候補のベクトルとを照合して、最もフィットすると判断された訳語をその都度採用している。このため、複数の訳語候補をもつことが多い技術／一般用語では、同じ原語であっても入力文の内容によって異なる訳語が採用される場合がある。つまり用語の訳ゆれは、「みんなの自動翻訳」に限らず現在のニューラル機械翻訳における共通の課題といえる。

【「専利」の訳し分けに関する課題】

「専利」は、中国における特許、実用新案、意匠の総称であり、対応する日本語が存在しない。このため本事業で作成した中国審決コーパスデータ中では、各案件の内容に応じて「特許」「実用新案」と訳し分けたが、当該コーパスを学習させた「みんなの自動翻訳」では「特許」と「実用新案」の訳し分け誤りによる誤訳が散見された。

「みんなの自動翻訳」で「専利」の訳し分けが発生したのは、中国審決コーパスデータを介して「特許」と「実用新案」という複数の訳語が学習されたためである⁴。学習データ中に複数の有力な訳語候補が存在することで、技術／一般用語と同様の訳ゆれが発生したものと考えられる。

こうした複数の訳語候補間の訳し分けを正確に行うには、ニューラル機械翻訳技術のさらなる技術発展が必要であろう。その間の対処策としては、中国審決コーパスデータから「専利」が「実用新案」と訳されている文を一次的に除外することで、「専利」の訳語を「特許」に固定することが可能である。本来「実用新案」と訳されるべき「専利」も一律「特許」と訳されてしまうことになるが、現状のランダム的な訳ゆれは解消する。

【長文における翻訳文の途絶】

「みんなの自動翻訳」における現状の課題としては、長文において機械翻訳文が最後まで出力されず、文の途中で途絶してしまう点も挙げられる。上限文字数など厳密な発生条件は特定不可能であったが、原文が極端な長文である場合、翻訳文の途絶が生じる可能性が高い。

機械翻訳では、このように長文において翻訳文の途絶が多発する場合、機械翻訳システムの性能、具体的にはバッファの上限を超えた結果である可能性が高い。この場合、ユーザが学習データ等を介して改善することは不可能である。

⁴ デフォルト状態の「みんなの自動翻訳」では「専利」は「特許」とのみ訳されており、「実用新案」と訳されるケースは存在しなかった。

ただし、本調査期間中も大規模アップデート以降、「みんなの自動翻訳」の長文処理能力は当初より明らかに改善されている。途絶は依然解消しないものの、各文とも出力される文字数が顕著に増加しており、アップデートを機にシステムのバッファが増強されたものと推察される。今後もハード・ソフト両面の技術発展に伴い、長文への対応もさらに改良されていくものと期待される。

【汎用 NT における日付誤り】

「みんなの自動翻訳」の現状の課題としてはもう一点、汎用 NT における、日付の日部分が異なる日付に変化する事象が挙げられる。原文中の日付の日部分が翻訳文中で異なる日付（多くの場合は 1 日）に誤訳され、本来の日付は文末に数値として出力される、という不備である。以下、実例を示す。

【文番号：#1078、汎用 NT】

原文	因此，本案的审查基础是专利权人在 2017 年 03 月 06 日提交的权利要求第 1-6 项。
機械翻訳文 第 20 回学習後	従って、本案の審理の基礎は特許権者が 2017 年 3 月 1 日に提出した請求項 1-6 である。06

本事象は第 19 回学習時から発生しており、評価用コーパスでは汎用 NT のみで発生している。日付を含む文でも発生するものとしなないものがあり、かつ、同一の文でも第 19 回時には発生せず第 20 回時のみで発生したものもあるなど、発生メカニズムは不詳である。

事象の内容や、一方のエンジンのみで第 19 回時から突如発生したという状況から、学習データである中国審決コーパスデータの問題とは考えにくく、エンジンのロジックに起因している可能性が高い。

[汎用 NT と特許 NT の優劣]

本調査では「みんなの自動翻訳」で利用可能な2つの翻訳エンジン「汎用 NT」と「特許 NT」の双方を対象とした。二者の違いは主にエンジン構築のための基礎学習データの違いであり、具体的には、特許 NT が特許文献由来の対訳コーパスのみを使用しているのに対し、汎用 NT は汎用の対訳コーパスと、特許 NT と共通の特許文献コーパスの一部とを併用している。無学習時はこの違いが顕著に表れており、例えば、特許文献由来のコーパスのみを重点的に学習した特許 NT では、汎用 NT では未知語扱いされた「技術方案」が「技術的解決手段」と正しく訳出されるなど、基本的な特許用語について一定の対応が見られた。

ただし、こうした基本的な特許用語を含めた特許・審決用語の大多数は中国審決コーパスデータの適用により汎用 NT でも早期に学習され、それ以降は、両者の翻訳品質に特段の差異は見いだせなくなった。エンジンの基礎学習の内容に依存する部分が大きい技術／一般用語に関しても、各文で頻繁に採用される有力な訳語候補は両エンジンほぼ同一で、例えば特定の訳語が一方のエンジンのみで頻繁に採用されるといった、基礎学習データに起因する顕著な差異は見られなかった。

自動評価指標 BLEU や RIBES、そして人手評価における内容伝達レベルなど、全体平均スコアでの比較による各評価結果も、学習初期には汎用 NT が特許 NT を若干上回ったものの、中期以降は両者の差異は誤差レベルの僅かなものとなっている。また、人手評価の一環として実施した「②2つの翻訳エンジンの相対評価」においても、学習を重ねるごとに個々の文単位でも両者同等の品質と評価される文が大勢を占めるようになっており、第20回時は100文中79文が両エンジン同等と評価された。

これらの状況から、中国審決コーパスデータを学習させる前提であれば、汎用 NT と特許 NT のどちらのエンジンを使用しても、得られる機械翻訳品質は同等であると結論できる。

ただし、上述のとおり、現「みんなの自動翻訳」では日付の日部分が異なる日付に変化する事象が発生しており、かつ、この事象は汎用 NT のみで発生している可能性が高い点に留意が必要である。

[第2回学習以降の評価スコアの停滞について]

自動評価指標 BLEU、RIBES の全体平均スコアは、第1回学習時に大幅な向上を見せたが、第2回以降は（第12回時を除き）微増と微減を繰り返すようになり、飽和とも映る様相を示した。人手による「内容伝達レベル」の評価も同様であり、第5回と第10回、第15回と第20回の品質は実質的に同等と見なせる。このことから、学習データは学習初期の数十万文対程度が適量と見なすこともできる。

ただし、一般に学習データ量の適量は機械翻訳エンジンにより異なり、また、同じ機械翻訳エンジンであってもバージョンによってその適量は変化する。本調査分析においても、第2回学習時以降の各回が横ばい傾向を見せるなか、「みんなの自動翻訳」の大規模アップデートが実施された第12回時のみは顕著なスコア向上を示し、以降の各回でもそのスコア水準が維持された。人手評価や実例の定点観測から、このときの品質改善は、学習データにおける出現頻度が訳語選定によりの確に反映されるようになったことで、不用意な訳語変化が減少したことが主な理由であることがわかった。ニューラル機械翻訳における訳語選定のメカニズムは、前述のとおり、学習データ中に存在する各訳語に周辺情報から多次元ベクトルを付与し、入力文の状況に応じて訳し分けるというものである。各訳語に付与されるベクトルは学習データ量が多いほど精密なものになるため、「みんなの自動翻訳」の大規模アップデートも、第12回当時の学習データ量でなかった場合、同等の改善効果が得られたとは限らない。

今後もソフト・ハード両面の技術発展に伴い、学習できるデータの規模は拡大していくと考えられる。将来の中国審決の機械翻訳の実用化に向けては、特定の機械翻訳システムにおける現状での飽和点にとらわれず、高精度の対訳コーパスを大量に作成して備えておくべきである。本事業では、現時点での機械翻訳システムに用いる学習データの整備という観点では十分に目的を達成したものの、機械翻訳技術の近年の発展は急速であり、その点は今後も注意が必要である。

また、本事業で作成した中国審決コーパスデータは、あくまで現在の中国審決の記載内容に対応するためのものである。中国の復審請求や無効宣告請求の増加が予測される中、審決の記載内容は、今後の法改正、法解釈を新たにする画期的な判決の登場、新たな技術分野への対応により、時代とともに変化していくことが予想される。特に2021年6月に施行された改正中国専利法には、懲罰的賠償の導入、法廷賠償金額の引き上げ、医薬品特許の存続期間延長制度の新設などが含まれ、当法改正により新たなタイプの審決文が増えることが予測される。このため、現在と同等の機械翻訳品質を長期にわたり維持するためには、今後も適切なタイミングで最新の中国審決のコーパスデータを適量作成し、これを追加学習させていくことが望ましい。

[本事業で達成された中国審決の機械翻訳品質について]

本調査では、本事業で作成した中国審決コーパスデータを機械翻訳システム「みんなの自動翻訳」に学習させることにより、中国審決の機械翻訳品質がどの程度改善されるかを検証した。その結果、全体としては「重要情報の 80%以上が正確に伝達される」レベルの機械翻訳精度が達成されることが確認された。

ニューラル機械翻訳では、特に定型的な文の翻訳精度が学習により高くなる。中国審決は、例えば審決の結論や適用条文など、定型的な文パターンで述べられる箇所が多く、今回の調査でも学習によって翻訳精度の大幅な改善が見られた。今回達成された翻訳品質レベルや、実際の機械翻訳結果の観測からも、少なくともこれら定型的な文が用いられる箇所に関しては、中国審決コーパスを学習させることにより、正確な翻訳文を提供することが可能であるといえる。

一方、全件学習後も文によっては誤訳等の不備が残存しているものもあり、また、同一案件内における技術／一般用語の訳ゆれなど、コーパスを学習させるだけでは解決が困難と思われる課題も検出された。しかし、現在の機械翻訳において 100%の精度を期待することは非現実的であり、機械翻訳サービスを利用するユーザもそのことは十分に承知している。

現在、中国審決情報は中国語でのみ提供されており、これに日本語で簡便にアクセスする手段は存在しない。本事業で作成した中国審決コーパスデータは、学習データとして「みんなの自動翻訳」に供することにより、中国審決を「重要情報の 80%以上が正確に伝達される」品質で日本語に機械翻訳することを可能とした。本コーパスを活用し、これを学習した機械翻訳エンジンで中国審決の和訳文を提供することで、中国審決への日本語によるアクセス性は大きく改善される。中国審決の内容を十分な精度で簡便かつ迅速に把握できるようになることで、我が国ユーザの中国における権利取得の業務の効率化に大いに貢献するものと考えられる。

I. 事業実施報告

1. 本事業の概要

本報告書は、2019年4月から2022年3月まで3か年にわたり実施した「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（以下「本事業」という。）」の実施内容の報告書である。本章において、本事業の概要についてまとめる。

1.1. 本事業の背景・目的

中国における特許・実用新案の出願件数の急増に伴い、中国での特許等の審判事件の処理件数も増加しており、我が国や欧米の処理件数を上回ってなお増加傾向にある。今後は我が国企業が関連する中国での審判事件が一層増加することが想定され、審決情報の重要性が高まっている。しかしながら、中国審決情報は一次情報としては中国語でのみ提供されており、我が国企業からは当該中国審決情報の日本語翻訳文の提供が強く要望されている。よって、我が国企業の円滑な権利取得と競争力向上の観点から、我が国企業による中国審決の分析や中国での審判事件への対応を効率化するため、中国審決情報の日本語への機械翻訳精度を向上させる必要がある。

現在、最も注目されている機械翻訳方式は、ニューラル機械翻訳である。ニューラル機械翻訳は、ニューラルネットワークによる深層学習技術を利用し、原文とその翻訳文とからなる対訳文のデータ（対訳コーパス。本報告書では単に「コーパス」ともいう。）を機械学習することにより翻訳を行うものである。従前の機械翻訳よりも高精度に自然な翻訳文を生成することができるが、専門性の高い文書を精度よく翻訳させるには、その文書に特有の用語や文体の学習が必要である。中国審決には専門的な審決用語や表現が多く含まれるため、高精度な機械翻訳文を提供するためには、実際の中国審決文に由来する対訳コーパスを十分な量作成し、翻訳エンジンに学習させる必要がある。

本事業では、中国審決を人手により翻訳して和訳文を提供するとともに、将来的な中国審決の機械翻訳文の提供に向け、和訳文から中国審決由来の対訳コーパスを作成する。また、ある用語を常に特定の訳語に翻訳できるようにするとともに、対訳コーパスを利用しない従来方式の機械翻訳へも対応すべく、対訳辞書も併せて作成する。これにより、我が国ユーザの中国審決に対するアクセス性を改善し、中国における権利取得の業務を効率化させることを目指す。

1.2. 本事業の内容

本事業では、中国審決全文を人手翻訳し、精度の高い和訳データ（「中国審決翻訳データ」）を作成するとともに、このデータを基にして、中国審決情報の機械翻訳精度の改善のための各種言語資源（「中国審決コーパスデータ」、「中国審決対訳辞書データ」等。第 1.3.項参照。）を作成した。

本事業ではさらに、実施期間中に作成した中国審決コーパスデータを機械翻訳システムに順次学習させ、学習の進行に伴う翻訳品質の改善の度合いや変化の傾向、実用にあたっての課題等について調査分析を行った。調査分析結果については第Ⅱ部で報告する。

1.3. 本事業の成果物

本事業で作成したデータは以下のとおりである。

1.3.1. 中国審決翻訳データ

中国国家知識産権局から入手した中国審決全文を人手で翻訳したデータである。翻訳対象とする中国審決は、重複文を排除した上で 200 万文、かつ、一文あたりの平均文字数が 70 文字を維持することを前提条件としたうえで、以下の優先事項により選定した。

- ・ 当事者が日本企業の案件を優先
- ・ 審決決定日が新しいものを優先
- ・ 分野の極度な偏りを避ける
- ・ 種別（無効審決と復審請求審決）の極度な偏りを避ける

実際の案件選定においては、概して文字数の少ない短文（見出し文、定型文が多い）ほど重複排除の対象となる確率が高いことから、重複排除後の平均文字数が 70 文字を大幅に超過する案件が多かった。また、分野や種別（無効／復審）による偏りも存在したが、こうした条件下、可能な限り上記優先事項を考慮し、全 19,182 件の中国審決を本事業の翻訳対象に選定した。以下、中国審決翻訳データの作成対象とした全 19,182 件の詳細を示す。

翻訳対象中国審決：19,182 件

総文数：2,694,693 文

総文字数：165,182,166 文字

重複排除後の総文数：2,000,136 文

重複排除後の総文字数：140,594,393 文字

重複排除後の平均文字数：70.3 文字

無効： 3,347 件（特許 1,152 件、実用新案 2,195 件）

復審： 15,835 件（特許 15,829 件、実用新案 6 件）

日本企業当事者案件：3,580 件

電気物理：9,203 件

化学：6,586 件

機械：3,393 件

※各審決の筆頭 IPC を調達仕様書別添 5 の表 1「各技術分野と国際特許分類（IPC）との対応関係」と照合して三分野に区分した。

2019 年度納品分： 6,324 件

2020 年度納品分： 7,094 件

2021 年度納品分： 5,764 件

合計：19,182 件

1.3.2. 中国審決コーパスデータ

中国審決の和訳である中国審決翻訳データ（⇒1.3.1.）とその原文である中国審決（中国語）から中日対訳文形式のデータを作成し、「中国審決コーパスデータ」とした。

中国審決コーパスデータは原則、中国審決翻訳データで翻訳対象とした全ての文について作成し、重複排除はしていない。したがって、コーパスの総文数は前項に示した翻訳対象中国審決の重複排除前の総文数とほぼ同数となるが、ごく少数、原文に重大な誤記があるなど対訳コーパスとしての精度を劣化させるおそれがある文については除外したため、若干の相違がある。

総文数：2,694,613 文

総文字数：165,175,803 文字

重複排除後の総文数：2,000,059 文

重複排除後の総文字数：140,588,296 文字

重複排除後の平均文字数：70.3 文字

電気物理：1,431,462 文

化学：891,155 文

機械：371,996 文

参考に、原文に重大な誤記があったため対訳コーパスから除外した文の実例を二例示す。

【審決番号：33997 無効】

原文	由此，在权利要求 1 具备创造性的情况下，直接或间接引用权利要求 1 的从属权利要求 2-9 同样不具备专利法第 22 条第 3 款规定的创造性。
翻訳文	よって，請求項 1 が創造性を有することから，請求項を直接又は間接的に引用する従属請求項 2 - 9 も同様に特許法第 22 条第 3 款に規定される創造性を有する。
除外理由	原文は前半が「具备 (=有する)」、後半が「不具备 (=有さない)」となっているが、両者は同じ語でなければ意味が通らず、一方が誤記と考えられる。翻訳時は前後の文脈から後者「不具备」を誤記と判断し、原文に沿わず「有する」と翻訳したが、これにより原文と翻訳文の内容が不一致となるため、学習データとしては不適切であり、悪影響の懸念があるため、コーパスからは除外した。

【審決番号：123318 復審】

原文	驳回决定中指出权利要求 1-14 相对于权利要求 1、2 和本领域公知常识的结合不具备专利法第 22 条第 3 款所规定的创造性。
翻訳文	拒絶査定 の 指摘 は 以下 の と お り で あ る : 請求項 1 - 14 は、引用文献 1、2 及び本分野の技術常識の組み合わせに対して特許法第 22 条第 3 款に規定される創造性を有さない。
除外理由	原文は「权利要求 1-14 (請求項 1 - 14)」の創造性を否定する根拠となる先行技術として「权利要求 1、2 (請求項 1, 2)」を示しており意味が通らない。翻訳者は誤記と判断し「引用文献 1、2」と訳したが、これにより原文と翻訳文が不一致となるため、コーパスからは除外した。

1.3.3. 中国審決対訳辞書データ

中国審決コーパスデータ (⇒1.3.2.) から名詞に相当する語句を抽出し、中国審決対訳辞書データを作成した。

具体的には、順次作成した中国審決コーパスデータに対し、言語処理技術を用いて名詞に相当する対訳語の候補を機械的に抽出し、過去事業で作成済みの語彙を除外したうえで、その時点での出現頻度の多い順に人手により採否判断を行った。そのうえで、コーパス全件作成完了後に改めて出現頻度をカウントし、上位の 10 万語を納品物とした。

1.3.4. 中国審決汎用文例リスト

中国審決に高頻度で出現する文とその翻訳文をリスト化した。具体的には、中国審決コーパスデータ全件における中国語原文の重複状況を調査したうえで、出現頻度が10回以上のもの5,412文を対訳形式のリストにまとめた。ただし、特許請求の範囲の摘記箇所など「公報由来⁵」の文は対象外としている。

出現頻度のカウントに際しては、日付や請求項番号などのバリエーションを集約すべく、原文・翻訳文とも数値部分は特定の記号{N}に置換した。リストにも数値部分{N}に置換したものを収録している⁶。

1.3.5. 中国審決翻訳データベース

中国審決翻訳データの作成対象となった中国審決19,182件（Wordファイル）の先頭2ページに掲載された各書誌事項と、各審決の本文冒頭部の「決定の要旨」の翻訳文を所定のフォーマットでデータ化し、特許庁内のデータベースに格納可能なデータに編集した。

⁵ 本事業では、コーパス各文に対し「公報由来」「審決由来」のラベリングを行った。「公報由来」文とは、実質的に文全体が特許公報等からの摘記である文を指す。

⁶ 特定の文における条文番号など、文の内容から数値部分が一意となる場合は数値のままとした。

II. 調查報告

2. 調査の概要

本事業の成果物である「中国審決コーパスデータ (⇒1.3.2.)」の機械翻訳に対する精度改善効果を検証するため、コーパスをニューラル機械翻訳システムに段階的に学習させ、中国審決の機械翻訳精度がどのように変化・向上していくか、及び機械翻訳文の変化の状況について調査分析した。第Ⅱ部では、本調査分析の内容及び結果について報告する。まず本章に調査分析の概要をまとめる。

2.1. 調査の方法

機械翻訳精度の調査では、自動評価指標 BLEU 及び RIBES による「自動評価」と (BLEU 及び RIBES の詳細については 3.2. を参照。)、評価者の目視による「人手評価」とを併用し、その評価結果に基づき機械翻訳精度を調査分析した。無学習の時点からコーパス全件学習後までの各学習段階の機械翻訳システムに対し、自動評価は全 21 回、人手評価は全 5 回を実施した。

機械翻訳の精度を評価するためには、翻訳対象とする中国審決文が必要となる。さらに、自動評価には、評価の基準となる「正解訳文」も必要である⁷。本事業では、事業初年度 (2019 年度) に初回バッチとして作成した中国審決コーパスデータから約 2,000 文対を「評価用コーパス」として選定し、上記用途に使用した。コーパス各文対の中国語原文を機械翻訳への入力文とし、それと対をなす人手翻訳文 (日本語) を自動評価のための正解訳文とした。なお、評価用コーパスとして選定した文は学習データには含めていない⁸。

本調査分析では、この評価用コーパスの各文を無学習時及び各学習段階の機械翻訳システムで翻訳し、その機械翻訳文に対して自動評価、人手評価の結果を取得して、学習の進捗に伴う翻訳品質の改善状況を調査した。さらに、各回の機械翻訳結果を定点観測し、機械翻訳文の変化の状況や誤訳の傾向、学習による改善点や改悪点などについて把握した。そのうえで、中国審決を機械翻訳する際の課題、その解決策等について考察した。

⁷ 第 3.2. 項を参照。

⁸ ただし、中国審決コーパスデータでは文の重複排除は行っていないため (⇒1.3.2.)、評価用コーパスと重複する文が学習データにも含まれることがある。

2.2. 使用した機械翻訳システム

本調査分析には、ニューラル方式の機械翻訳システム「みんなの自動翻訳@TexTra®」⁹を用いた。具体的には、同システムの提供する、学習タイプの異なる2種の翻訳エンジン「汎用NT」及び「特許NT」¹⁰を併用した。両エンジンともデフォルト（無学習）状態を起点として、学習データである中国審決コーパスデータを約10万文ずつ段階的に学習させ¹¹、各回の学習前後の機械翻訳結果の内容や品質評価値を比較した。

2.2.1. 汎用NTと特許NT

「みんなの自動翻訳」の「汎用NT」エンジンと「特許NT」エンジンとは、エンジン作成に使用したプログラム自体は同一であるが、エンジン構築のために使用した基礎的な学習データがそれぞれ異なる。具体的には、特許NTでは特許文献由来の対訳コーパスのみを使用しているのに対し、汎用NTでは汎用の対訳コーパスと、特許対訳コーパスの一部とを併用している。また、翻訳処理の前後に施される前処理、後処理も両者で細部が異なる。

2.2.2. 「みんなの自動翻訳」のバージョンアップについて

「みんなの自動翻訳」は定期的にバージョンアップが行われる。本調査分析期間中も全7回のバージョンアップが実施された。バージョンアップは、各回の機械翻訳処理時に得られるバージョン情報により把握できる。

バージョンアップに伴い旧バージョンは使用不可能となるため、本調査分析では常にその時点で最新のバージョンのエンジンを使用した。以下、調査期間中に実施されたバージョンアップの一覧を示す。

表 2.2.1. 「みんなの自動翻訳（汎用NT／特許NT）」バージョンアップ一覧

バージョン情報	対象学習回
GPMT-3.8_190531_nmt	無学習時～第1回学習時
GPMT-3.8_190930_nmt	第2回学習時～第3回学習時
GPMT-3.8_200131_nmt	第4回学習時～第8回学習時
GPMT-3.8_200531_nmt	第9回学習時～第11回学習時

⁹ <https://mt-auto-minhon-mlt.ucri.jgn-x.jp/content/menu>

¹⁰ これらは2022年3月現在のエンジンの名称であり、本事業開始当初は別名称の旧エンジン（汎用NMT、特許NMT）を用いていた。本報告書では混乱を避けるため、全編にわたり汎用NT、特許NTで統一した。

¹¹ 厳密には、「みんなの自動翻訳」では1回学習させたエンジンに対して再度追加の学習を行うことは不可能であるため、学習データを増量する都度、デフォルト状態のエンジンに対して過去分を全件一括で学習させ直す形となる。

GPMT-3.9_200930_nmt	第 12 回学習時～第 13 回学習時
GPMT-3.9_210131_nmt	第 14 回学習時～第 16 回学習時
GPMT-3.9_210531_nmt	第 17 回学習時～第 18 回学習時
GPMT-3.9_210930_nmt	第 19 回学習時～第 20 回学習時

2.2.3. 「みんなの自動翻訳」の大規模アップデート

「みんなの自動翻訳」のバージョン情報は、例えば「GPMT-3.8_190531_nmt」のように 2 つの数値を含むが、定期的なバージョンアップでは後半部の 6 桁の日付部分のみが更新される。これに対し、第 12 回学習前に実施されたバージョンアップは、前表のとおり日付部分に加えて前半の数値も 3.8 から 3.9 に更新されており、通常より上位のバージョンアップであることがわかる。

通常の定期的なバージョンアップでは、その前後で機械翻訳の傾向に特段の違いは見られなかったが、この第 12 回学習前のバージョンアップのみは、その後の機械翻訳文の品質や訳語選定の傾向に明白な変化が生じた。本報告書でもその影響について各所で言及することとなるため、本報告書では第 12 回学習前のバージョンアップのみ、他の定期的なバージョンアップと区別して「『みんなの自動翻訳』の大規模アップデート」と称する。

2.3. 自動評価と人手評価のタイミング

BLEU 及び RIBES による自動評価は、本事業で作成した中国審決コーパスデータを約 10 万文ずつ段階的に学習させる都度、具体的には第 1 回から第 20 回までの各回学習後に実施した。一方、人手評価はコーパスを約 50 万文ずつ段階的に学習させる都度、つまり第 5 回学習後、第 10 回学習後、第 15 回学習後、第 20 回学習後に実施した。なお、自動評価、人手評価とも、これらに加えて無学習時（デフォルト状態）の翻訳エンジンによる機械翻訳結果も評価している。

表 2.3.1. 自動評価と人手評価の実施タイミング

学習回数	無学習	1 回	2 回	3 回	4 回	5 回	6 回	7 回	8 回	9 回	10 回
自動評価	○	○	○	○	○	○	○	○	○	○	○
人手評価	○	—	—	—	—	○	—	—	—	—	○
		11 回	12 回	13 回	14 回	15 回	16 回	17 回	18 回	19 回	20 回
		○	○	○	○	○	○	○	○	○	○
		—	—	—	—	○	—	—	—	—	○

2.4. 評価用コーパス

調査分析に用いる評価用コーパスとして、事業初年度初回バッチで作成した中国審決コーパスデータから、以下の方針に基づき 2,013 文を選定した。

2.4.1. 案件単位の選定

本事業では中国審決を案件単位で全文翻訳したが(⇒1.3.1.)、評価用コーパスも案件単位での選定を基本方針とした。これは、同一の案件に属する文が評価用コーパスと学習データの双方に振り分けられることを避けるためである¹²。

中国審決コーパスデータを学習させた機械翻訳システムを実用する際、翻訳対象となる中国審決はコーパスに含まれない新規の案件が想定される。このため本調査分析も極力この状況をシミュレートして行う必要がある。同一の審決に由来する各文は共通の内容、用語、表現を多く含むため、その一部を評価用コーパスとし、残りを学習データに含めた場合、機械翻訳上不当に有利な条件となり、実用時の機械翻訳品質を正確に推定できない。

上記理由から、評価用コーパスは案件単位でその全文を選定する方針とし、全 31 案件を選定した。ただし、これら 31 案件間での重複文については一文のみ選定し、他は選定対象外とした。選定対象外となった重複文は初回の学習データに含まれることになるが、案件間で重複する文は中国審決全体に頻出する定型文と見なされ、学習データ中にこうした文が既に存在している状況は実用時においても十分に想定されるため、不問とした。

¹² 本事業で作成した中国審決コーパスデータは、評価用コーパスに採用された文以外は全て学習データとして使用している。

2.4.2. 復審請求／無効宣告請求について

中国審決には「復審請求審決」と「無効宣告請求審決」の二種が存在する。中国審決全体では前者が圧倒的多数であり、中国審決コーパスデータにおいても同様の状況であるが（⇒1.3.1.）、評価用コーパスは双方が同等の案件数となるように選定した（復審：16 案件、無効：15 案件）。

また、審決の決定内容（結論）についても、復審請求審決においては「拒絶査定維持（6 案件）」と「拒絶査定撤回・再審査（10 案件）」の両パターン、無効宣告請求審決においては「権利維持（7 案件）」「全部無効（5 案件）」「一部無効（3 案件）」の全パターンを含めた。

2.4.3. 法区分について

中国審決の法区分は「特許」を中心に選定したが（29 案件）、「実用新案」も 2 案件含めた。なお、中国では実用新案の実体審査は原則行われなため、復審請求審決もほとんど発生しない¹³。このため実用新案として選定した 2 案件はいずれも無効宣告請求審決となった。

2.4.4. 技術分野について

電気物理分野：14 案件、化学分野：11 案件、機械分野：6 案件とした。この比率は本事業で翻訳対象とした中国審決全体の分野別構成比率とおおむね合致している。

2.4.5. 適用条文について

中国審決には、該審決が依拠した中国特許法条文を示す「適用条文」が明示されている。審決で論じられる内容は、同種（復審請求／無効宣告請求）の審決であっても適用条文により大きく異なる。反面、適用条文が同一の審決では、審決に至る考察の内容や論理の流れ、使用される用語・表現に一定の類似性がある。

中国審決には、多くの案件で頻繁に適用される特許法条文がいくつか存在する。評価用コーパスには、最頻出の適用条文として、創造性に関する「第 22 条第 3 款」、新規性に関する「第 22 条第 2 款」、明細書のサポートに関する「第 26 条第 4 款」、記載の明確性に関する「実施細則第 20 条第 1 款」、補正が可能な範囲に関する「第 33 条」のそれぞれに該当する案件を中心に選定した。さらに、さほど頻出でない適用条文の翻訳精度も把握するため、「第 2 条（技術的手段を構成）」、「実施細則第 72 条（無効宣告請求取下げ）」が適用された案件も加えた。

¹³ 本事業の翻訳対象には実用新案案件の復審請求審決が 6 件含まれる（⇒1.3.1.）。このうち 4 件は方式審査で拒絶されたもの、2 件は実体審査で拒絶されたものである。後者の実体審査は、実施細則第 44 条第 1 項第 2 号に基づく強化方式審査を指すものと推察される。

2.4.6. 一部案件における「特許請求の範囲」部分の除外

評価用コーパス選定の基本方針は項番 2.4.1. に記載したとおり案件単位であるが、限られた文数で極力多数の案件／パターンを網羅するため、特許公報等に由来し（つまり審決特有の文ではなく）、かつ、文数が比較的多くなる「特許請求の範囲」の摘記部分については、約半数の案件（復審：8 案件、無効：8 案件）で、評価用コーパスの選定から除外した。

除外した各摘記文は学習用データに含まれることとなるため、同じ案件で評価用コーパスに採用された文と同一の技術用語や表現が学習データ側に確実に含まれてしまうというシミュレーション上のデメリットが生じるが、その反面、より多くの案件を評価用コーパスに含められるというメリットも得られる。本調査分析ではこのメリットを優先する方針とした。

2.4.7. 選定結果

上記の各方針に基づき、全 31 案件、2,013 文を「評価用コーパス」として選定した。文単位では復審：1,043 文、無効：970 文であり、平均文字数は 76.96 文字である。

2.4.8. 人手評価対象文について

本調査分析では、品質評価として自動評価と人手評価を定められたタイミングで実施した（⇒2.1.）。自動評価は評価用コーパス全件（2,013 文）を対象とするが、人手評価については、評価用コーパスのうちの 100 文を選定して精密に実施した。

「人手評価対象文」の詳細については、第 4 章「人手評価結果とその分析」の第 4.4.項「人手評価の対象文」において述べる。

3. 自動評価結果とその分析

3.1. 自動評価の概要

自動評価は、「みんなの自動翻訳」が提供する2種の機械翻訳エンジン「汎用 NT」及び「特許 NT」(⇒2.2.) について、それぞれデフォルト（無学習）状態での機械翻訳結果に対する1回と、200万文からなる中国審決コーパスデータを約10万文ずつ段階的に学習させた都度の機械翻訳結果に対する全20回、合計21回にわたり実施した。使用した自動評価指標は BLEU と RIBES の二種類である。

3.2. BLEU と RIBES

本調査分析で自動評価に用いた二種類の指標 BLEU¹⁴及び RIBES¹⁵はいずれも、あらかじめ与えられた「正解訳文」を評価の基準とし、評価対象の機械翻訳文がこの正解訳文とどの程度一致しているかに基づいて 0.0000（最低スコア）～1.0000（最高スコア）の範囲内の評価スコアで示す仕組みである。

本調査分析では中国審決コーパスデータから 2,013 文を評価用コーパスとして選定し、各文対の中国語原文を機械翻訳への入力文としたが、自動評価においては、それと対をなす日本語文（本事業で作成した人手翻訳文）を「正解訳文」として使用した。

¹⁴ Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu : BLEU : a Method for Automatic Evaluation of Machine Translation, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 311–318, 2002.

¹⁵ 平尾努、磯崎秀樹、須藤克仁、Duh Kevin、塚田元、永田昌明：語順の相関に基づく機械翻訳の自動評価法、自然言語処理、Vol.21, No.3, pp.421–444, 2014.

図 3.2.1. BLEU、RIBES の自動評価スコア算出方法の概要

評価用コーパス (文番号#1)		機械翻訳結果
原文 (機械翻訳入力文)	人手翻訳文 (正解訳文)	(特許 NT 第 20 回学習後)
<p>如果所属技术领域的技术人员根据说明书公开的内容, 结合本领域公知常识, 能够容易地将发明扩展到该权利要求的保护范围, 并且也没有证据表明该发明在该权利要求的范围内不可以实施, 那么, 该权利要求的保护范围能得到说明书的支持。</p>	<p>当業者が明細書に開示された内容に基づき, 本分野の技術常識を組み合わせ, 発明を該請求項の保護範囲に容易に拡張することができ, 且つ本発明の該請求項の保護範囲内で実施できないという証拠も示されていないならば, 該請求項の保護範囲は明細書にサポートされている。</p>	<p>当業者が明細書に開示されている内容に基づき, 本分野の技術常識を組み合わせ, 発明を該請求項の保護範囲に容易に拡張することができ, 且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合, 該請求項の保護範囲は明細書にサポートされている。</p>

評価用コーパス原文の「機械翻訳結果」を「正解訳文 (評価用コーパスの人手翻訳文)」と比較し、その一致度に基づき評価スコアを自動的に算出する。

BLEU:0.7313、RIBES : 0.9463

3.3. 各回の自動評価スコアの集計結果

無学習時から第 20 回学習後まで全 21 回実施した自動評価の結果を本項にまとめる。

3.3.1. 評価用コーパス全件の平均スコア

評価対象とした 2 つの翻訳エンジン（汎用 NT と特許 NT）それぞれについて、各回学習後における評価用コーパス全件の BLEU 及び RIBES の平均スコアの推移を下二表及びグラフに示す。

表 3.3.1. 評価用コーパス全件の BLEU 平均スコア一覧

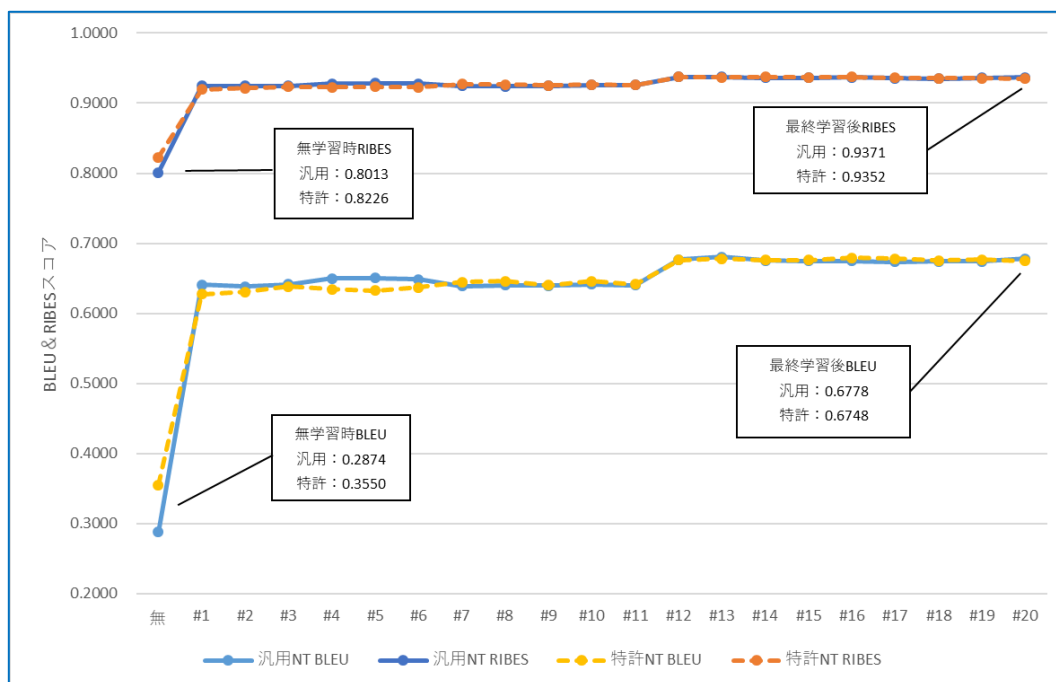
学習回数	学習データ量 ¹⁶	バージョン	汎用 NT		特許 NT	
無学習時	0	3.8_190531	0.2874	前回比	0.3550	前回比
第 1 回学習後	128,376	〃	0.6406	+0.3532	0.6273	+0.2723
第 2 回学習後	260,048	3.8_190930	0.6379	-0.0027	0.6302	+0.0028
第 3 回学習後	393,769	〃	0.6417	+0.0038	0.6381	+0.0080
第 4 回学習後	531,457	3.8_200131	0.6498	+0.0081	0.6346	-0.0035
第 5 回学習後	670,229	〃	0.6507	+0.0009	0.6327	-0.0019
第 6 回学習後	805,167	〃	0.6486	-0.0021	0.6370	+0.0042
第 7 回学習後	937,950	〃	0.6391	-0.0095	0.6447	+0.0077
第 8 回学習後	1,070,296	〃	0.6402	+0.0011	0.6462	+0.0015
第 9 回学習後	1,205,286	3.8_200531	0.6394	-0.0008	0.6401	-0.0061
第 10 回学習後	1,340,277	〃	0.6417	+0.0023	0.6456	+0.0055
第 11 回学習後	1,473,544	〃	0.6403	-0.0014	0.6414	-0.0042
第 12 回学習後	1,606,773	3.9_200930	0.6766	+0.0363	0.6762	+0.0347
第 13 回学習後	1,750,160	〃	0.6804	+0.0038	0.6778	+0.0016
第 14 回学習後	1,885,334	3.9_210131	0.6752	-0.0052	0.6757	-0.0022
第 15 回学習後	2,020,165	〃	0.6747	-0.0005	0.6761	+0.0004
第 16 回学習後	2,156,082	〃	0.6750	+0.0003	0.6791	+0.0030
第 17 回学習後	2,290,129	3.9_210531	0.6734	-0.0016	0.6782	-0.0009
第 18 回学習後	2,425,849	〃	0.6742	+0.0008	0.6752	-0.0030
第 19 回学習後	2,559,464	3.9_210930	0.6743	+0.0001	0.6764	+0.0012
第 20 回学習後	2,692,560	〃	0.6778	+0.0035	0.6748	-0.0016
無学習時と第 20 回学習後のスコア差			+0.3904		+0.3198	

¹⁶ 「学習データ量」の欄にはその時点で学習データに供された中国審決コーパス全件（累計文数）を掲載した。各回で増量される学習データは重複文を排除すると約 10 万文となるが、コーパス自体は重複排除せず使用するため（⇒1.3.2.）、3 割ほど文数が多くなっている。

表 3.3.2. 評価用コーパス全件の RIBES 平均スコア一覧

学習回数	学習データ量	バージョン	汎用 NT		特許 NT	
無学習時	0	3.8_190531	0.8013	前回比	0.8226	前回比
第 1 回学習後	128,376	〃	0.9247	+0.1234	0.9198	+0.0972
第 2 回学習後	260,048	3.8_190930	0.9244	-0.0003	0.9214	+0.0016
第 3 回学習後	393,769	〃	0.9248	+0.0004	0.9237	+0.0023
第 4 回学習後	531,457	3.8_200131	0.9278	+0.0029	0.9226	-0.0011
第 5 回学習後	670,229	〃	0.9283	+0.0005	0.9235	+0.0008
第 6 回学習後	805,167	〃	0.9278	-0.0005	0.9230	-0.0005
第 7 回学習後	937,950	〃	0.9245	-0.0033	0.9271	+0.0041
第 8 回学習後	1,070,296	〃	0.9239	-0.0006	0.9265	-0.0006
第 9 回学習後	1,205,286	3.8_200531	0.9246	+0.0007	0.9252	-0.0013
第 10 回学習後	1,340,277	〃	0.9260	+0.0014	0.9264	+0.0012
第 11 回学習後	1,473,544	〃	0.9262	+0.0002	0.9257	-0.0008
第 12 回学習後	1,606,773	3.9_200930	0.9374	+0.0112	0.9372	+0.0115
第 13 回学習後	1,750,160	〃	0.9378	+0.0004	0.9371	-0.0001
第 14 回学習後	1,885,334	3.9_210131	0.9364	-0.0013	0.9372	-0.0027
第 15 回学習後	2,020,165	〃	0.9363	-0.0001	0.9367	-0.0006
第 16 回学習後	2,156,082	〃	0.9367	+0.0004	0.9374	+0.0007
第 17 回学習後	2,290,129	3.9_210531	0.9353	-0.0014	0.9365	-0.0009
第 18 回学習後	2,425,849	〃	0.9350	-0.0003	0.9356	-0.0009
第 19 回学習後	2,559,464	3.9_210930	0.9363	+0.0013	0.9357	+0.0001
第 20 回学習後	2,692,560	〃	0.9371	+0.0008	0.9352	-0.0005
無学習時と第 20 回学習後のスコア差			+0.1358		+0.1126	

図 3.3.1. 評価用コーパス全件の BLEU、RIBES 平均スコア推移



上掲のとおり、双方のエンジン（汎用 NT 及び特許 NT）、双方の評価指標（BLEU 及び RIBES）とも、中国審決コーパスデータの学習の結果、スコアは大幅に向上した。無学習時から最終学習後にかけて全期間トータルのスコアの向上度を下表に再掲する。

表 3.3.3. 無学習時と最終学習後の BLEU、RIBES スコア比較

学習回		無学習	第 20 回	向上度
汎用 NT	BLEU	0.2874	0.6778	+0.3904
	RIBES	0.8013	0.9371	+0.1358
特許 NT	BLEU	0.3550	0.6748	+0.3198
	RIBES	0.8226	0.9352	+0.1126

各学習段階のスコアの変動を見ると、学習データは約 10 万文ずつ均等に増量されたのに対し、スコアの向上は初回学習時に大きく偏った。例えば汎用 NT の初回学習後の全体平均スコアは BLEU で前回比+0.3532、RIBES では+0.1234 と、いずれも上表に示した全期間トータルの向上度の 90%強を占めている。特許 NT も同様に、初回学習後のスコア向上度は BLEU が+0.2723、RIBES が+0.0972 と、それぞれ全期間トータルの向上度の 85%強を占めている。

一方、第 2 回学習後以降のスコアは各回とも微動にとどまり、グラフからも見て取れるように、事実上の横ばい傾向が続いた。初回学習後から第 11 回学習後までのスコアの向上度は汎用 NT で BLEU : **-0.0003**、RIBES : **+0.0015**、特許 NT で BLEU : **+0.0141**、RIBES : **+0.0059** と、学習データの総量が初回時の 128,376 文から第 11 回時には 1,473,544 文と 10 倍以上に増量されていることを考えると、ごく小規模な変動にとどまっている。

こうした中、第 12 回学習後のみは汎用 NT で BLEU が **+0.0363**、RIBES が **+0.0112**、特許 NT で BLEU が **+0.0347**、RIBES が **+0.0115** と、それまでの 10 回の学習でのトータルを大きく上回る、初回時以来の明らかなスコア向上が見られた。これは主に、第 12 回学習前に実施された「みんなの自動翻訳」の大規模アップデート (⇒2.2.3.) の影響と見られる。「みんなの自動翻訳」の大規模アップデートによる翻訳品質への影響に関しては、第 3.4.項で行う機械翻訳実例の定点観測や、第 4 章にまとめた人手評価結果において詳しく分析・考察する。

その後、第 13 回後以降は、第 12 回後に上昇したスコア水準を保ちつつ、再び微増・微減を繰り返す横ばい状態となった。第 12 回後と最終第 20 回後とのスコア比較では、汎用 NT では BLEU が **+0.0012**、RIBES が **-0.0003**、特許 NT では BLEU が **-0.0014**、RIBES が **-0.0020** といずれも誤差レベルの微小な変動にとどまっており、第 12 回後に向上したスコア水準が一時的なものではなく、最終回後まで安定して維持されたことがわかる。

二種のエンジン、すなわち汎用 NT と特許 NT のスコアを比較すると、無学習時は特許 NT が有意に高いスコアを得ていたが、中国審決コーパスデータ学習後は両エンジンともおおむね同水準のスコアで推移しており、品質上の優劣が無くなったことが示された。両エンジンには基礎学習データの内容が異なり (⇒2.2.1.)、無学習時には特許文献由来のコーパスのみを用いて構築された特許 NT の方に何らかのアドバンテージが存在したが、そのアドバンテージは中国審決コーパスデータの学習を通じて汎用 NT でも早期に獲得され、以降は両エンジンの基礎学習データの違いは機械翻訳品質に大きな影響を及ぼさなくなったという推移が見て取れる。

以上、各学習回での自動評価の全件平均スコアの推移を概観した。汎用 NT、特許 NT とも、中国審決コーパスデータの学習によって BLEU、RIBES 双方のスコアが顕著に向上しており、本事業で作成したコーパスが中国審決文に対して高い翻訳品質改善効果を有していることが示された。

ただし、コーパスの学習がもたらした機械翻訳文の具体的な改善内容や、改善後の翻訳精度が本事業の目的の達成に足る水準に達しているかについては自動評価スコアのみからは

確言できないため、機械翻訳文の実例を用いた検証や人手評価の集計結果の分析によりさらに検証していく。

【付記】

評価用コーパスには2,013文を選定したが(⇒2.4.)、このうちの19文は調査分析開始当時の「みんなの自動翻訳」では文長制限に抵触し機械翻訳文が取得できなかった。このため本項に示した全体平均スコアは、全期間を通じてこれら19文を除いた1,994文を対象としている。翻訳不能文19文の詳細は第3.4.6.項を参照されたい。

3.3.2. 文単位のスコア向上／低下カウント結果

前項に示した各回学習後の自動評価スコアの推移は評価用コーパス全件（1,994 件）の平均スコアの推移であり、評価用コーパス中の個々の文が一律に平均スコアと同じ推移を示しているとは限らない。むしろ、学習後に全体平均スコアが向上していても、個々の文レベルではスコアが低下したものや、スコアが変化しなかったものも少なからず存在すると考えられる。この点に鑑み、本項では、各回の学習前後で BLEU 及び RIBES の評価スコアが向上した文、変化しなかった文、低下した文をそれぞれカウントした。以下、集計結果を示す。

表 3.3.4. 文単位 BLEU/RIBES スコア向上／不変／低下カウント結果

	汎用 NT						特許 NT					
	BLEU			RIBES			BLEU			RIBES		
	向上	不変	低下	向上	不変	低下	向上	不変	低下	向上	不変	低下
第 1 回	1,912	16	66	1,754	17	223	1,863	16	115	1,711	23	260
第 2 回	730	482	782	757	489	748	882	335	777	866	355	773
第 3 回	816	477	701	794	485	715	871	352	771	878	361	755
第 4 回	823	513	658	795	521	678	754	437	803	791	442	761
第 5 回	723	530	741	742	536	716	767	427	800	764	436	794
第 6 回	722	527	745	702	546	746	847	394	753	808	413	773
第 7 回	693	442	859	696	455	843	858	459	677	857	488	649
第 8 回	805	445	744	796	455	743	775	470	749	744	481	769
第 9 回	740	493	761	752	502	740	714	454	826	704	460	830
第 10 回	783	465	746	795	472	727	803	483	708	792	482	720
第 11 回	730	486	778	718	494	782	735	466	793	732	475	787
第 12 回	1,031	468	495	990	479	525	1,066	392	536	1,055	405	534
第 13 回	689	702	603	656	735	603	628	767	599	661	783	550
第 14 回	619	634	741	637	642	715	646	668	680	631	683	680
第 15 回	602	779	613	600	788	606	586	787	621	572	803	619
第 16 回	600	778	616	642	757	595	616	784	594	645	774	575
第 17 回	660	625	709	617	636	741	638	652	704	609	682	703
第 18 回	633	747	614	609	758	627	597	804	593	576	810	608
第 19 回	685	589	720	712	604	678	634	707	653	612	725	657
第 20 回	701	701	592	665	717	612	566	873	555	563	877	554

自動評価スコアの変動は、機械翻訳文の変化とほぼ同義である。まれに機械翻訳文が変化してもスコアが同値となる場合はあるが（このため BLEU と RIBES の「不変」文数は各回とも異なる）、少なくともスコアが変動した機械翻訳文は、学習前後で何らかの変化が生じていると確定できる。

この視点で前表を見ると、まず初回学習後においては、無学習時からスコアが「不変」であった文は両エンジン、両指標とも全体の 1%程度であり、ほぼ全ての機械翻訳文に何らかの変化が生じている。大多数の文はスコアが「向上」しており、前項における初回学習後に全体平均スコアが大幅に上昇という状況と整合する。

一方、それ以降の推移を見ると、第 2 回～第 11 回学習後にかけては、各回とも「不変」文は 400～500 文前後で、残る 1,500～1,600 文は「向上」文と「低下」文がおおむね半々という状況が続いている。つまりこの期間は、学習の都度、全体の 4 分の 3 の機械翻訳文に何らかの変化が生じ続けていたことになる。この期間の全体平均スコアは毎回微動にとどまり横ばい傾向であったが、本カウント結果から、その理由は各文の機械翻訳結果がほとんど変化しなくなったからではなく、大半の機械翻訳文は変化し続けていたが、スコアが向上した文と低下した文とが毎回ほぼ均衡していたためであったことがわかる。

この傾向が変化したのが、「みんなの自動翻訳」の大規模アップデート（⇒2.2.3.）が実施された直後の第 12 回学習後である。第 12 回後のカウント結果は、「不変」文の数はほぼ変わらないものの「向上」文と「低下」文のバランスが前者に大きく偏り、後者の約 2 倍となっている。この回は全体平均スコアも顕著に向上しており、文単位のカウントもこれと同期した変動を示している。さらに、第 13 回学習後以降は「不変」文の比率が明らかに増加し、大規模アップデート以前は各回 400～500 文程度であったところ、第 13 回後以降は各回 600～800 文前後を占めるようになった。この調査結果から、「みんなの自動翻訳」における学習の都度、大半の翻訳文が変化し続けるという傾向が、大規模アップデートによってやや抑制されたことがわかる¹⁷。

とはいえ、第 13 回学習以降においても全体の過半数の機械翻訳文は毎回変化しており、「向上」と「低下」がおおむね均衡している点も大筋では変わらない。大規模アップデート以降は若干抑制されたとはいえ、学習の進捗にかかわらず大半の機械翻訳文が変化し続け、かつ、スコアの向上と低下がほぼ同程度となるという状況は「みんなの自動翻訳」の本質的な傾向として最終学習後まで不変であった。

¹⁷ 大規模アップデートのタイミングは第 12 回学習前であるが、大規模アップデート後の機械翻訳結果同士の間隔は第 13 回時からとなる。

この状況は、文単位の自動評価スコアの増減が、額面どおり実際の翻訳品質の向上／低下を示していると仮定した場合、学習データの増量の都度、全体の約 3 分の 1 の文において品質が向上するが、その反面、同じく約 3 分の 1 の文では品質が低下するということになる。つまり本調査結果は、第 2 回学習以降の中国審決コーパスデータの追加の都度、学習効果とほぼ同程度の副作用が生じている可能性があることを示している。この点に関しては、第 3.4.項で行う機械翻訳文実例の定点観測や第 4 章の人手評価結果の分析時に詳しく検証する。

3.3.3. 審決種別、適用条文別、当事者別の集計結果

評価用コーパスは、第 2.4.項で述べたとおり、中国審決のバリエーションを幅広くカバーするように選定している。本項では、自動評価スコアを各種バリエーション別に集計した結果を示す。なお、比較を容易にするため、集計結果は無学習時と最終学習後のみを表示する。

3.3.3.1. 審決種別ごとの集計

中国審決は大きく「復審請求」と「無効宣告請求」に分けられる。さらに無効宣告請求には特許を対象としたものと実用新案を対象としたものがある（復審請求は特許のみ）。本項ではこれら審決の種別ごとに自動評価スコアを集計・比較した。以下、結果を示す。

表 3.3.5. 審決種別ごとの自動評価スコア比較

審決種別	文数	NT	BLEU			RIBES		
			無学習	第 20 回	向上	無学習	第 20 回	向上
復審 (特許)	1,034	汎	0.2776	0.6936	+0.4160	0.7997	0.9420	+0.1423
		特	0.3810	0.6912	+0.3102	0.8332	0.9404	+0.1072
無効宣告 (特許)	861	汎	0.2942	0.6515	+0.3573	0.8006	0.9295	+0.1289
		特	0.3259	0.6480	+0.3221	0.8082	0.9271	+0.1189
無効宣告 (実用)	99	汎	0.3301	0.7419	+0.4118	0.8244	0.9527	+0.1283
		特	0.3358	0.7371	+0.4013	0.8373	0.9511	+0.1138
全件	1,994	汎	0.2874	0.6778	+0.3904	0.8013	0.9371	+0.1358
		特	0.3550	0.6748	+0.3198	0.8226	0.9352	+0.1126

中国審決コーパスデータでは全 19,177 案件のうち 15,831 案件を「復審（特許）」が占めており、これに比べて「無効宣告（特許）」は 1,151 案件、「同（実用）」は 2,195 案件と少数である（⇒1.3.1.）。つまり「無効宣告（特許）」と「同（実用）」は、学習データ中の構成比が「復審（特許）」に比べて小さい。この学習データ量の差によって、各審決種別に学習効果の差が生じている可能性がある。

学習による品質改善効果は、最終学習後に到達したスコア水準（上表「第 20 回」欄）と、無学習時から最終学習後にかけてのスコアの向上度（上表「向上」欄）とから総合的に判断するのが妥当である。

まず最終学習後のスコアを見ると、汎用 NT、特許 NT とも、学習データが最も少量であった「無効宣告（特許）」の BLEU スコアが他の 2 種よりも有意に低い。ただし、それに次いで少量な「無効宣告（実用）」では、両エンジンともに最終学習後の BLEU、RIBES スコアの双方で相対順位が最高位となっており、全体的には各種別のスコア順位は学習データ

量の多寡と比例していない。

スコアの向上度に関しては、BLEU ではやはり「無効宣告 (特許)」の向上度がやや低い傾向にあり、汎用 NT では三種中最下位、特許 NT でも最下位とほぼ同水準の第 2 位であったが、一方の RIBES では「無効宣告 (特許)」は僅差ながら汎用 NT で第 2 位、特許 NT では第 1 位とむしろ上位に属しており、やはり三種の審査種別の状況に、学習データ量の多寡との明瞭な関連性は見いだせない。

バリエーション別のスコア集計は、全件平均スコアと異なり、各種別それぞれ異なる文のスコアを集計・比較している。このため各種別の文集合間の翻訳難易度等の条件は一律ではなく、ある程度の誤差が生じることは避け難い。このことを考慮すると、本集計結果はむしろ、学習データ量の顕著な差にもかかわらず、各審決種別がおおむね同等の品質向上を遂げていると見るべきである。

つまり本集計結果からは、学習データ量の少なかった「無効宣告 (特許)」及び「同 (実用)」も、学習データ量が圧倒的に多い「復審 (特許)」と遜色のない翻訳改善効果が得られている可能性が高いと結論できる。

3.3.3.2. 適用条文による集計

中国審決では、審決の基礎となる特許法条文が「適用条文」として示されている。条文によって多数の審決で頻繁に適用されるものとまれにしか適用されないものがあり、中国審決コーパスデータにおいても適用条文ごとの構成比にばらつきが生じている。

評価用コーパスには、第 2.4.5.項で述べたとおり、マイナーな適用条文として「特許法第 2 条第 2 款 (2 案件)」、「実施細則第 72 条 (1 案件)」を含めている。本項では、これらマイナーな条文が適用された審決と、それ以外のメジャーな条文が適用された審決 (28 案件) の自動評価スコアを比較した。下表に結果を示す。

表 3.3.6. 適用条文ごとの自動評価スコア比較

審決種別	文数	NT	BLEU			RIBES		
			無学習	第 20 回	向上	無学習	第 20 回	向上
特許法 第 2 条	132	汎	0.2589	0.6575	+0.3986	0.7915	0.9348	+0.1433
		特	0.4276	0.6476	+0.2200	0.8527	0.9294	+0.0767
実施細則 第 72 条	39	汎	0.3444	0.6446	+0.3002	0.8430	0.9280	+0.0850
		特	0.3155	0.6543	+0.3388	0.7975	0.9345	+0.1370
それ以外	1,823	汎	0.2882	0.6800	+0.3918	0.8011	0.9375	+0.1364
		特	0.3506	0.6772	+0.3266	0.8210	0.9356	+0.1146
全件	1,994	汎	0.2874	0.6778	+0.3904	0.8013	0.9371	+0.1358
		特	0.3550	0.6748	+0.3198	0.8226	0.9352	+0.1126

マイナーな条文である「特許法第 2 条」「実施細則第 72 条」とも、最終学習後の BLEU スコアは両エンジンでメジャーな条文より低くなっており、学習データ量の多寡と比例しているように映る。ただし、一方の RIBES スコアでは両条文とも双方のエンジンでメジャーな条文とおおむね同水準といえる。また、スコア向上度も、各条文、両エンジンとも項目によってはメジャーな条文を上回っているなど、総合的には学習データ量の多寡が集計結果に顕著に表れているとはいえない。

前項で述べたとおり、バリエーション別の集計結果の比較では一定の誤差が不可避である。特に適用条文別の集計の場合、評価用コーパスを案件単位で選定しているため、「特許法第 2 条」は 2 案件 (132 文)、「実施細則第 72 条」は 1 案件 (39 文) のみの集計値となっている。このため、案件ごとの難易度のばらつきがほとんどならされていない¹⁸。

¹⁸ 第 2 条に属する 2 案件の間にも、最終学習後の BLEU スコアは案件#1 (47793 復審) が汎 : 0.6282、特 : 0.6330 なのに対し、案件#2 (77695 復審) は汎 : 0.6877、特 : 0.6627 と大きな差異がある。

こうした条件を考慮すると、本集計結果における条文間のスコア差は誤差の範囲として扱うのが妥当であろう。つまり本調査結果からは、マイナーな適用条文の審決であっても、中国審決コーパスデータ全件の学習により、他のメジャーな適用条文の審決とほぼ遜色ない翻訳品質改善効果が得られていると見なせる。

3.3.3.3. 技術分野間の差異

最後に、評価用コーパスを技術分野ごと、具体的には化学（C）、電気物理（E&P）、機械（M）の三分野に細分化して集計した。下表にその結果を示す。

表 3.3.7. 技術分野ごとの自動評価スコア比較

審決種別	文数	NT	BLEU			RIBES		
			無学習	第 20 回	向上	無学習	第 20 回	向上
化学	683	汎	0.2997	0.6835	+0.3838	0.8100	0.9397	+0.1297
		特	0.3571	0.6773	+0.3202	0.8247	0.9378	+0.1131
電気物理	922	汎	0.2759	0.6875	+0.4116	0.7928	0.9379	+0.1451
		特	0.3537	0.6851	+0.3314	0.8217	0.9365	+0.1148
機械	389	汎	0.2930	0.6448	+0.3518	0.8061	0.9307	+0.1246
		特	0.3543	0.6459	+0.2916	0.8211	0.9276	+0.1065
全件	1,994	汎	0.2874	0.6778	+0.3904	0.8013	0.9371	+0.1358
		特	0.3550	0.6748	+0.3198	0.8226	0.9352	+0.1126

中国審決には、技術用語・表現も多数含まれ、かつ、あらゆる技術分野が対象となり得る。このため、技術用語全般への対応度も機械翻訳品質に大きく影響する。一方、中国審決コーパスデータは、技術用語も含むとはいえ、そのカバー範囲は限定的・非網羅的である。このため本調査における技術用語の翻訳品質は、機械翻訳システムのデフォルトでの学習内容（基礎学習）に委ねられる部分が多い。したがって、本項に示した技術分野ごとの集計結果は、「みんなの自動翻訳」のデフォルトでの各技術分野の翻訳品質に偏りがいないか、顕著に品質が劣る分野が存在しないかを確認することが主な目的となる。また、基礎学習データの異なる 2 種のエンジン間の技術用語への対応度に差異があるか否かも調査する必要がある。

始めに、基礎学習データの異なる汎用 NT と特許 NT とで技術用語の翻訳品質に有意な差があるかを考察する。上掲した自動評価スコアを比較すると、各技術分野とも無学習時は特許 NT のほうがスコアが有意に高くなっている。第 2.2.1.項に記したとおり、特許 NT の基礎学習に用いられた特許由来の学習データの一部は汎用 NT では使用されておらず、こ

の違いが技術用語の翻訳品質差を生んでいるようにも映る。

だが、仮にそうであった場合、中国審決コーパスデータがカバーする技術用語は限定的であることから、両エンジン間のスコア差は容易に縮まらないものと考えられる。しかしながら、実際には、両エンジンの全体平均スコアはコーパス初回学習後には早々に同水準となり（⇒3.3.1.）、以降も両エンジン間に実質的なスコア差は生じていない。上掲した最終学習後（第20回）のスコアを見ても、各分野ともエンジン間に有意な差はない。この推移から、無学習時における両エンジンのスコア差は技術用語の翻訳品質の差によるものではなく、中国審決コーパスデータの学習によって速やかに改善される性質のもの、すなわち特許・審決用語への対応度の差に起因しているものと考えられる。換言すれば、特許・審決用語への対応度の差が解消された最終学習後の自動評価スコアが各技術分野とも両エンジン同等であったことから、両エンジンの各分野の技術用語の翻訳品質に、基礎学習データの違いによる顕著な差異は生じていないと結論される。

この結論を前提に、「みんなの自動翻訳」における技術分野別の翻訳品質差の有無について考察する。各技術分野で集計される文はそれぞれ異なる文であり、各文に含まれる技術用語以外の要素の翻訳難易度もさまざまであるが、仮にこうした要素の翻訳難易度の差を不問とすれば、上表に示した各分野間のスコア差が「みんなの自動翻訳」による技術用語の翻訳品質の差と見なせる。その観点で各分野のスコアを見ると、無学習時、最終学習後とも、それぞれのエンジンでの各分野のスコア水準はおおむね同等であり、強いて挙げれば両エンジンとも機械分野のスコアが他の2分野に比して低いが、スコアが極端に劣る分野は存在しない。

異なる文集合同士の比較による誤差を考えれば、この集計結果からは「みんなの自動翻訳」のデフォルト学習内容に技術分野間の顕著な偏りはなく、各技術分野を均等にカバーしていることがうかがえる。

ただし、「みんなの自動翻訳」における技術用語の翻訳品質が実用に足るレベルかは、自動評価スコアからは直接判断できない。そもそも上記スコアは技術用語のみの評価結果ではなく、仮にスコアが低調であったにせよ、それが技術用語の問題なのか、他の理由によるものかはスコアからは判別できない。この点に関しては、次章に示す人手評価結果などから改めて分析・判断する必要がある。

3.4. 実例を用いた機械翻訳結果の定点観測

本項では、評価用コーパスから実例をいくつかピックアップし、各回学習後の機械翻訳文の変化の状況を定点観測する。学習の各段階における機械翻訳文の変化の状況を追跡し、中国審決コーパスデータが「みんなの自動翻訳」にもたらした具体的な学習効果の内容や、機械翻訳品質の実質的な改善の度合いについて確認する。また、自動評価の全体平均スコアから看取された翻訳品質の推移の検証や、各実例に付された自動評価スコアと実際の翻訳品質との照合も併せて行う。

3.4.1. 実例 3-1：文番号#1、特許 NT

最初の実例として、評価用コーパスの文番号#1 を取り上げる。まずは本例の原文及び正解訳文、そして特許 NT による無学習時及び第 1 回学習後の機械翻訳結果を下欄に示す。

【文番号：#1、特許 NT】

原文	如果所属技术领域的技术人员根据说明书公开的内容，结合本领域公知常识，能够容易地将发明扩展到该权利要求的保护范围，并且也没有证据表明该发明在该权利要求的范围内不可以实施，那么，该权利要求的保护范围能得到说明书的支持。	
正解訳文	当業者が明細書に開示された内容に基づき、本分野の技術常識を組み合わせ、発明を該請求項の保護範囲に容易に拡張することができ、且つ本発明の該請求項の保護範囲内で実施できないという証拠も示されていないならば、該請求項の保護範囲は明細書にサポートされている。	
無学習	当業者が明細書に開示された内容に基づいて、本分野の一般的な知識と組み合わせて、本発明を該特許請求の保護範囲に容易に拡張することができ、且つ該発明が該特許請求の範囲内で実施できないことを示唆するものではなく、そのため、該特許請求の保護範囲は明細書のサポートを得ることができる。	
	—	BLEU：0.4415 (－) RIBES：0.8494 (－)
第 1 回学習後	当業者が明細書で開示されている内容に基づき、本分野の技術常識を組み合わせ、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	良化	BLEU：0.6971 (+0.2556) RIBES：0.9426 (+0.0932)

上掲の一段目が機械翻訳の入力文となった「原文」であり、二段目が自動評価の基準となる「正解訳文」、すなわち本事業で作成した人手翻訳文である。この 1～2 段目の対訳文が中国審決コーパスデータの 1 文対に相当する。そして三段目には無学習時の機械翻訳文、四

段目には第 1 回学習後の機械翻訳文をそれぞれ表示している。本例では、これらの機械翻訳文で学習の前後で変化した箇所を青字・赤字で示した。正解訳文と同じ訳である場合は青字、異なる訳である場合は赤字としている。また、各回の機械翻訳文の末行には目視による品質評価（前回時より**良化**／**同等**／**悪化**）と、実際に付された BLEU と RIBES のスコアを記載した。

まず無学習時の機械翻訳結果を見ると、例えば「当業者」や「**明細書のサポート**」といった、やや専門性の高い特許用語も適訳語で訳されている。これは、特許 NT が基礎学習に特許文献由来のデータを用いたことによるアドバンテージと見られる。

ただし、その一方で、原文には単に「**发明**」としか書かれていないにもかかわらず「**本發明**」と不要なニュアンスが追加されていたり、「請求項」を意味する「**权利要求**」を「**特許請求**」と直訳していたりと、基本的な特許用語に対応できていない箇所もある。また、本例は文全体が中国審決の頻出表現で構成されているが、翻訳文後半の「**示唆するものではなく、そのため**」という接続や、文末の「**得ることができる**」という訳語も、正解訳文と対比すると明らかのように、本来の文意から外れており、こうした審決頻出表現への対応も不十分である。

これに対し、中国審決コーパスデータ第 1 回学習後の機械翻訳結果では、「**权利要求**」→「**請求項**」、「**一般的な知識**」→「**技術常識**」など、全ての特許用語が正解訳文のとおりに出されるようになった。「**本發明**」、「**示唆するものではなく、そのため**」、「**得ることができる**」といった特許・審決表現の誤訳も全て解消しており、原文の内容が正確に把握できる翻訳文に改善されている。

本例に見られる初回学習後の翻訳品質の顕著な改善は、第 3.3.1.項における自動評価の全体平均スコアの初回学習後の大幅な向上と一致している。このことは、本例に見られる特許・審決用語や審決頻出表現の顕著な改善が、同じタイミングで評価用コーパス全体にわたり発生したことを示唆している。

なお、本件に付された自動評価スコア自体も全体平均同様、BLEU、RIBES とともに大幅な向上を示している。目視による実質的な品質評価も大幅な「**良化**」であり、学習前後の相対評価としては両指標とも正当な評価となっている。

ここで、両エンジンの比較のため「汎用 NT」の機械翻訳結果も確認しておく。

【文番号：#1、汎用 NT】

無学習	本技術分野の技術者が説明書に開示された内容に基づいて、本分野の常識を踏まえて、発明を当該権利の保護範囲に拡張することができ、この権利要求の範囲内で実施できないことを証明する証拠がない場合、その権利要求の保護範囲は説明書の支持を受けることができる。	
	—	BLEU：0.3009（－） RIBES：0.8030（－）
第1回学習後	当業者が明細書に開示された内容に基づき、本分野の技術常識を組み合わせ、該発明を該請求項の保護範囲内に容易に拡張することができ、また該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	良化	BLEU：0.7253（+0.2556） RIBES：0.9300（+0.0932）

汎用 NT の無学習時の機械翻訳結果では、特許 NT で正しく訳せていた「当業者」が「本技術分野の技術者」、「明細書」が「説明書」、「サポートされる」が「支持を受ける」と誤訳されているのを始め、「請求項」も「権利」や「権利要求」と訳されるなど、文中のほぼ全ての特許用語が直訳調の不適訳となっている。汎用 NT の基礎学習にも特許文献由来のコーパスは一部使用されているとはいえ、この結果を見る限り、無学習時の特許用語への対応度は特許 NT のほうが優れている。

ただし、翻訳文の構文自体はむしろ汎用 NT のほうが正確であり、例えば特許 NT で「本発明」「示唆するものではなく、そのため」と不備が発生していた箇所はそれぞれ「発明」「示す証拠もない場合」と正しい解釈で訳されている（文末の「受けることができる」は特許 NT と同様に誤訳）。

このため無学習時の翻訳文はどちらのエンジンが優れているか一概には言えないが、自動評価スコアは BLEU、RIBES とともに特許 NT のほうが高くなっている。自動評価スコアは正解訳文との一致度に基づいて算出されるため、特許用語の一致度の差が強く反映されたものと推察される。第 3.3.1.項に示した全体平均スコアでも、無学習時点の汎用 NT は特許 NT よりも明らかにスコアが劣っており、本例と同様「特許用語への対応度の差」がその主な原因となっていると考えられる。

これに対し、汎用 NT の第 1 回学習後の機械翻訳文では、上記全ての特許用語が特許 NT と同じく正解訳文どおりの適訳語に訳されるようになった。つまり本例では、初回学習後の特許用語の翻訳品質は両エンジン完全に同等となっている。この状況も、汎用 NT の初回学習後の全件平均スコアが特許 NT と同等の水準に達したという推移と整合している。つまり本例からは、無学習時の両エンジンの優劣の原因であった（基本的な）特許用語への対応

度の差は、中国審決コーパスデータの初回学習により、大多数の文では早々に解消されたという状況が見て取れる。

一方、特許用語以外の部分では初回学習後の翻訳文でも両エンジン間に微妙な相違が生じており、汎用 NT では「保護範囲内に」という無学習時には発生していなかった新たな誤訳も生じている。コーパス学習後に発生した以上、初回分のコーパスが誘発したエラーということになるが、同じ学習データを用いている特許 NT ではこの誤訳は発生していない。

上記以外にも両エンジン間で微妙な相違はあるが（例えば「明細書で開示されている内容」と「明細書に開示された内容」）、いずれも同義の範囲内の変化であって、品質上の優劣はない。このため両者の優劣は上記「保護範囲内に」の誤訳の有無のみであり、この誤訳が生じていない特許 NT のほうが優れている。これに対し、汎用 NT の翻訳文に付された BLEU スコア (0.7253) は特許 NT (0.6971) のものよりも高く、実態とは異なる。

続いて、再び特許 NT の第 2 回学習後の機械翻訳文を示す。

【文番号：#1、特許 NT】

第 2 回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識と組み合わせ、発明を該請求項の保護範囲内に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	悪化	BLEU : 0.6967 (-0.0004) RIBES : 0.9286 (-0.0140)

第 2 回学習後の翻訳文でも、全ての特許・審決用語は前回と同じ適訳語が安定して採用されており、文全体を通して前回からほとんど変化は見られない。微細な変化は 3 箇所で見られるが、このうち 2 箇所は「明細書で開示されている」→「明細書に開示されている」及び「技術常識を組み合わせる」→「技術常識と組み合わせる」という、内容理解上どちらでも優劣のない助詞の変化である。

ただし、3 箇所目の変化にあたる「保護範囲内に」は、汎用 NT で初回学習後に発生したのと同じ誤訳であり、この誤訳の発生により今回の翻訳品質は前回よりも「悪化」している。

ここで、汎用 NT の第 2 回学習後の機械翻訳文を示す。

【文番号：#1、汎用 NT】

第 2 回学習後	当業者が明細書に開示された内容に基づき、本分野の技術常識と組み合わせて、該請求項の保護範囲に 発明 を容易に拡張することができ、 且つ 該発明が該請求項の範囲内で 実施することができないことを証明する 証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	良化	BLEU：0.6663 (-0.0590) RIBES：0.9295 (-0.0005)

こちらも特許 NT と同様、各所に同義の範囲内の微細な変化が生じているが、前回時に発生していた「保護範囲内に」の誤訳は今回解消されている。

つまり第 2 回学習後では、両エンジンに同一の学習データを用いたにもかかわらず、特許 NT では「保護範囲内に」の誤訳が新たに発生し、汎用 NT では前回発生していた同じ誤訳が今回は解消されるという、正反対の変化が生じたことになる。前は汎用 NT のみで誤訳が発生していたという経緯を含めて考えると、両エンジン共通の学習データ（基礎学習データも含む）中に「保護範囲内に」という訳語候補が初回学習の時点から存在しており、それが何らかの理由で双方のエンジンでランダム的に採用されている状況と映る。

第 2 回学習後の特許 NT の実質的な翻訳品質に関しては、上記「保護範囲内に」という誤訳の発生により、前回時よりも「悪化」と判定した。自動評価結果も前回よりスコアを下げっており、増減の方向としては実際の品質変化と合致している。ただし、特に BLEU に関しては、誤訳の重大度に比べてスコアの下げ幅は誤差レベルの僅かなものであった。一方、汎用 NT では誤訳が解消したことで実際の翻訳品質は「良化」しているのだが、自動評価では BLEU、RIBES とともにスコアが低下しており、引き続き実態と乖離してしまっている。

続いて、第 3 回～第 5 回学習後の特許 NT の機械翻訳結果について列挙する。

【文番号：#1、特許 NT】

第 3 回学習後	当業者が明細書に 開示された 内容に基づき、本分野の技術常識を 組み合わせて 、 発明 を該請求項の保護範囲 まで 容易に拡張することができる場合、該発明が該請求項の範囲内で 実施できないことを示す証拠もなく、よって 、該請求項の保護範囲は明細書にサポートされている。	
	悪化	BLEU：0.6901 (-0.0066) RIBES：0.9350 (+0.0064)

第4回学習後	当業者が明細書に開示された内容に基づき、本分野の技術常識を組み合わせ、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施されないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	良化	BLEU : 0.7417 (+0.0516) RIBES : 0.9426 (+0.0076)
第5回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識を組み合わせることで、発明を該請求項の保護範囲に拡張することが容易であり、且つ該発明が該請求項の範囲内で実施できないことを示す証拠がない場合、該請求項の保護範囲は明細書にサポートされているといえる。	
	同等	BLEU : 0.6300 (-0.1117) RIBES : 0.9228 (-0.0198)

第3回から第5回学習後にかけても、特許・審決用語は各回全ての語が正解訳文どおりの適訳語で非常に安定している。一方、一般用語・表現に関しては、「開示された内容」→「開示されている内容」や「組み合わせて」→「組み合わせることで」、「実施できない」→「実施されない」など、同義表現内での不安定な変化が引き続き各回各所で発生している。例えば「開示された内容」と「開示されている内容」は、前者が無学習時と第3回、第4回後、後者が第1回、第2回と第5回後と、毎回この二つの訳語候補からランダムに選ばれているかのように採用されている。なお、第2回後に発生していた「保護範囲内」の誤訳は第3回時には解消され、以降発生していない。

この「保護範囲内」の誤訳を含め、本例各所の一般用語・表現に見られるランダム的な訳語変化は、「みんなの自動翻訳」が採用するニューラル翻訳方式での訳語選定のメカニズムに起因するものと見られる。「みんなの自動翻訳」でのこのメカニズムの詳細は非開示であるが、以下、一般的なニューラル翻訳の訳語選定のメカニズムを簡略化して説明する。

ニューラル翻訳では、学習データである大量の対訳文対から翻訳モデルを構築するという点では従来の統計翻訳方式と同様であるが、統計翻訳のようにフレーズテーブル（句単位の対訳辞書）を生成・利用するのではなく、学習データに基づいて個々の単語や構文を多次元のベクトルで表現している¹⁹。具体的には、学習データ中の各文例の各用語を、その用語の周辺に存在する語句と関連付けて学習し、多次元ベクトル化する。そして学習データ中の全ての文例、用語のベクトルを集約して「翻訳モデル」を作成する。さらに、作成した翻訳モデルを用いて学習データ中の各原文を仮翻訳し、これと学習データ中で原文と対をなす

¹⁹ 江原暉将、岡俊行：ニューラル機械翻訳における訳語誤りについての分析，2019.

https://japio.or.jp/00yearbook/files/2019book/19_4_05.pdf

訳文とを比較する。翻訳モデルは学習データ全体を集約したものであるため、個々の文レベルでは仮翻訳結果と学習データ中の訳文とは異なることが多くなる。この場合、仮翻訳結果が訳文に近づくようベクトルを調整し、翻訳モデルを更新する。この翻訳モデルの更新と仮翻訳とを繰り返して、仮翻訳結果と学習データの訳文の差が収束したら、その時点で最も仮翻訳結果との差の小さい翻訳モデルを採用する。これが一般的なニューラル翻訳における学習の手順である。

こうした学習メカニズムにより、ニューラル機械翻訳では、ごく平易に述べると、ある用語について学習データ中に複数の訳語が存在する場合、それぞれの訳語に対し、その訳語が使われた各文例における周辺情報に基づいた多次元ベクトルが付与される。そして、訳語選定の際は、入力文における用語の周辺情報と、学習データ中の各訳語候補のベクトルとを照合して、最もフィットすると判断された訳語が採用される。

この多次元ベクトルは、各用語間の学習データ全体における共起関係に基づくため、学習データが追加されると、その追加データ中に存在しない訳語であっても、周辺に位置する語のベクトルの変化により、間接的に変化する可能性がある。学習の都度、このように文中の各語のベクトルが変化するため、同じ入力文であっても学習前後で採用される訳語や文構造が変化する場合がある。

特許・審決用語の場合、その多くは中国審決コーパスデータで統一訳語が定められており、事実上、学習データ中に他の訳語候補が存在しない。このため学習の都度ベクトルが変化しても、ほかに選択の余地がないため、常時安定してコーパスどおりの適訳語が採用されていると考えられる。これに対し、本例各所に見られるような一般用語・表現は学習データ中に複数の訳語候補が存在し、かつ、各回で追加されるデータに、周辺語が異なる新たな文例が含まれる可能性が高いため、学習前後で訳語のベクトルが変化する傾向が強い。特に「みんなの自動翻訳」の場合、このベクトルが学習データ追加の都度大きく変化する傾向があると見られ、これにより訳語間の優先順位が頻繁に入れ替わっているものと推察される。これが外部からは、ランダム的な訳語変化と映る。

本例を見る限りは、「みんなの自動翻訳」のランダム的な訳語変化はおおむね同義の範囲内に収まっているようだが、「保護範囲内に」など不適訳が突発的に採用されてしまうこともあるようである。例えばこの期間においても、第5回学習後のみ文末に「といえる」という不要な文言が採用されている。

第3回～第5回学習後の実質的な翻訳品質に関しては、第3回後は、前回悪化の原因となった「保護範囲内に」という誤訳は解消されたが、さらに重大な「拡張することができる

場合、…証拠もなく、よって、」という誤訳が新たに発生しており、前回よりさらに「悪化」している。この誤訳は第4回後には解消され、「実施されない」というやや不適な訳語が採用された以外はほぼ完璧な内容に戻ったため「良化」、そして第5回後は、「実施されない」は「実施できない」と元に戻ったものの、前記のとおり文末の不要な「といえる」など軽微な改悪もあり、差し引きで「同等」と判定した。したがって、第3回～第5回にかけての実質的な翻訳品質は「悪化」→「良化」→「同等」という推移となる。これに対し、BLEUスコアの変動は「悪化」→「(大幅な)良化」→「(さらに大幅な)悪化」、RIBESは「良化」→「良化」→「(大幅な)悪化」をそれぞれ示しており、実際の品質との不整合が散見されるとともに、評価指標間でも矛盾が生じている。

続いて、第6回学習後から第10回学習後までの機械翻訳結果を示す。

【文番号：#1、特許 NT】

第6回学習後	当業者が明細書に開示されている内容に基づいて、本分野の技術常識と組み合わせ、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	良化	BLEU：0.6804 (+0.0504) RIBES：0.9371 (+0.0143)
第7回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識を組み合わせ、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	同等	BLEU：0.7313 (+0.0509) RIBES：0.9463 (+0.0032)
第8回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識と組み合わせ、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内に実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	悪化	BLEU：0.6805 (-0.0508) RIBES：0.9388 (-0.0075)
第9回学習後	当業者が明細書に開示された内容に基づいて、本分野の技術常識と組み合わせ、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	良化	BLEU：0.7079 (+0.0274) RIBES：0.9426 (+0.0038)

第 10 回学習後	当業者が明細書に開示された内容に基づき、本分野の技術常識を組み合わせ、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠がない場合、該請求項の保護範囲は明細書にサポートされている。	
	同等	BLEU : 0.7561 (+0.0482) RIBES : 0.9481 (+0.0055)

第 6 回学習後以降も、特許・審決用語は各語とも常時安定している。一方、一般用語・表現に関しては、第 5 回後に発生していた軽微な不備（文末の不要な「といえる」など）も解消され、第 8 回後における「範囲内に実施できない」という助詞の突発的な改悪を除けば、各回とも「同義語の範囲内でのランダム的な訳語変化」にとどまっている。この期間の実質的な品質変化はこの一箇所の誤訳のみであり、翻訳品質が安定してきた。この状況は、実質的に現状維持状態であったこの期間の全体平均スコアの推移とおおむね合致している。

ただし、本例単独の自動評価スコアでは、特に BLEU スコアで毎回大幅な変動が繰り返されており、依然として実際の翻訳品質の推移に即していない。その点、RIBES の変動は比較的小規模であり、もともと BLEU より変動幅が小さい性質ではあるが、こちらのほうが実態に近い印象を受ける。

続いて、第 11 回学習後から第 15 回学習後までの機械翻訳結果を示す。

第 11 回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識を組み合わせ、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠がない場合、該請求項の保護範囲は明細書にサポートされている。	
	同等	BLEU : 0.7257 (-0.0304) RIBES : 0.9426 (-0.0055)
第 12 回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識を組み合わせ、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	同等	BLEU : 0.7313 (+0.0056) RIBES : 0.9463 (+0.0037)
第 13 回学習後	当業者が明細書に開示された内容に基づき、本分野の技術常識を組み合わせ、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	同等	BLEU : 0.7618 (+0.0305) RIBES : 0.9518 (+0.0055)

第 14 回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識を組み合わせて、発明を該請求項の保護範囲に容易に 拡張でき 、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	同等	BLEU : 0.6758 (-0.0860) RIBES : 0.9396 (-0.0122)
第 15 回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識を組み合わせて、発明を該請求項の保護範囲に容易に 拡張することができ 、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。	
	同等	BLEU : 0.7313 (+0.0555) RIBES : 0.9463 (+0.0067)

第 11 回～第 15 回学習後の機械翻訳文も傾向はこれまでと変わらず、特許・審決用語が毎回安定して同じ適訳語に訳される一方、一般用語・表現はこれまでと同様の箇所が、同様の訳語候補間でランダム的に変化している。唯一、これまで常時安定して「拡張することができ」と訳出されていた箇所が第 14 回学習後のみ「**拡張でき**」に変化した。が、「一般用語・表現の同義の範囲内での変化」という意味では他と同様である。結果、各回の品質評価は全て「同等」となった。

第 3.3.1.項で述べたとおり、「みんなの自動翻訳」では第 12 回学習前に大規模なアップデートが行われており、同じタイミングで自動評価の全体平均スコアが顕著に向上した。第 3.3.2.項の文単位のスコア増減カウントでも第 12 回学習後はスコア「向上」文が大幅に増加しており、大規模アップデートを機に何らかの品質改善が広範囲に生じたものと考えられる。しかしながら、本例に関しては第 12 回後の翻訳文は前回から「証拠がない」が「証拠もない」へと 1 文字変化したのみで、「同等」との評価結果が示すとおり、アップデート前後の翻訳品質に特段の変化は見られない。付された BLEU、RIBES スコアも、増加はしているが誤差レベルの微増である。

ただし、本例の場合、アップデート以前からほぼ正確な翻訳文が得られており、これ以上改善の余地がない状況であった。このため大規模アップデートの好影響が具体的な変化として顕出しなかったものと推察される。大規模アップデート時の品質改善効果については、後続の実例で引き続き検証していく。

最後に、第 16 回学習後から第 20 回（最終学習回）学習後までの機械翻訳結果を示す。

第 16 回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識を組み合わせて、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。
	同等 BLEU : 0.7313 (±0) RIBES : 0.9463 (±0)
第 17 回学習後	当業者が明細書に開示されている内容に基づいて、本分野の技術常識を組み合わせて、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。
	同等 BLEU : 0.7011 (-0.0302) RIBES : 0.9408 (-0.0055)
第 18 回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識を組み合わせることで、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。
	同等 BLEU : 0.7185 (+0.0174) RIBES : 0.9408 (±0)
第 19 回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識を組み合わせて、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。
	同等 BLEU : 0.7313 (+0.0128) RIBES : 0.9463 (+0.0055)
第 20 回学習後	当業者が明細書に開示されている内容に基づき、本分野の技術常識を組み合わせて、発明を該請求項の保護範囲に容易に拡張することができ、且つ該発明が該請求項の範囲内で実施できないことを示す証拠もない場合、該請求項の保護範囲は明細書にサポートされている。
	同等 BLEU : 0.7313 (±0) RIBES : 0.9463 (±0)

第 16 回から第 20 回学習後までにかけても、特許・審決用語は常時安定して適訳語に訳され、一般用語・表現に関しても、前半部の「基づき」と「基づいて」、及び「組み合わせ」と「組み合わせることで」の二箇所が、第 17 回後から第 19 回後にかけて二つの訳語候補間で変化したのみであった。いずれも同義の範囲内の変化であり、実質的な翻訳品質は各回とも完全に同等である。

なお、第 3.3.2.項では、大規模アップデート以降の傾向の変化として、学習前後でスコアが「不変」の文、すなわち機械翻訳文が変化しなかった（と見なせる）文の比率が明らかに増加したことも指摘した。本例においても、第 16 回後、第 20 回後は前回と完全一致の翻

訳文が出力されており、その他の回の翻訳文も、学習前後で変化は生じているとはいえ、その変化は大規模アップデート以降、明らかに小規模・局所的になってきている。

先述のとおり、本例に見られる一般用語や文構造のランダム的な変化は、ニューラル機械翻訳の訳語選定のメカニズムに起因しており、学習データの追加によって各訳語に付与されるベクトルが広範に変化するためであると考えられる。このため本例に限らずあらゆる文で同様に発生し、各学習回における機械翻訳文の広範囲な変化をもたらしていた。大規模アップデート以降この傾向がやや抑制されたのは、「みんなの自動翻訳」において、ベクトルの算出ロジックが見直された結果である可能性が高い。

本例でも、「保護範囲内」や「拡張することができる場合, …証拠もなく, よって,」「実施されない」、「といえる」、「範囲内に実施できない」など、アップデート以前には突発的に採用されていた不適訳語や不適な文構造の多くが、アップデート以降は一切採用されなくなった。本例を見る限り、こうした不適訳語は訳語候補から事実上外され、例えば「開示されている」と「開示された」、「基づいて」と「基づき」といった少数の有力かつ同義の候補の間での変化に限定されるよう、ベクトルの算出ロジックが改良されたように見受けられる。具体的にどのような改良が施されたかは外部からはうかがい知れないが、少なくとも各回で採用される訳語の精度が改善され、ランダム的な訳語変化が生じにくくなったことは確かである。「不変」文の増加も、各文においてランダム的な訳語変化が抑制された結果、一部の文では翻訳文が全く変化しなかったためであろう。

換言すれば、大規模アップデート以降の「みんなの自動翻訳」では、以前に比べて学習データの内容がよりの確に訳語選定に反映されるようになり、学習データ中で使用頻度の高い有力な訳語候補が順当に採用されることが多くなったといえるであろう。その結果、各文において不用意な訳語変化、すなわち不適訳語の突発的な採用が抑制されたことが、大規模アップデート時の自動評価スコアの顕著な向上をもたらした、少なくとも一因であると考えられる。

以上、評価用コーパスから文番号#1 をピックアップして、無学習時から最終学習後まで、特許 NT による各回の機械翻訳文の内容を定点観測した。一例のみの分析であるが、少なくとも本例からは、以下の傾向が見て取れた。

- ・ 特許・審決用語に関しては、無学習時においては特許 NT で基本的な用語の一部に対応できていたが、汎用 NT ではほとんど対応できていなかった。これに対し、中国審決コーパスデータの初回学習後は、両エンジンとも全ての特許・審決用語が正解訳文どおりの適訳語に改善され、以降も常時この適訳語で完璧な安定を見せた。

自動評価の全体平均スコアが初回学習後に大幅に向上し、以降もその水準が保たれていることから、他の文においても本例と同様に特許・審決用語が初回学習によって正解訳文どおりの適訳語に一挙に改善し、そのまま安定したものと推察される。

- ・ 無学習時における特許 NT と汎用 NT との品質差は、主に特許用語への対応度の差によるものであった。中国審決コーパスデータを介した特許・審決用語の学習により、特許 NT の優位性は早期に失われ、以降、両エンジンの品質は同水準となった。

- ・ 一般用語・表現や文構造に関しては、特許・審決用語と異なり、学習の都度、不安定に変化する傾向が顕著であった。こうした一般用語や文構造のランダム的な変化は、「みんなの自動翻訳」が採用するニューラル翻訳方式における訳語選定のメカニズムと、学習データ中に複数の訳語候補が存在するという一般用語の性質とによるものと推察される。

本例を見る限り、こうした一般用語や文構造の変化の多くは同義の範囲にとどまり、実質的な翻訳品質には影響しない。一部、品質悪化を招く不適訳語への突発的な変化も見られたが、このような場合も次回学習後には再び変化して適訳語に戻るが多かった。

- ・ 第 3.3.2.項に示した文単位のスコア向上／低下カウント結果では、特許・審決用語が安定した第 2 回学習後以降も全体の 4 分の 3 の機械翻訳文が変化し続けた。上述した「みんなの自動翻訳」における一般用語や文構造のランダム的な変化が主な原因と考えられる。なお、このように多数の文が変化を繰り返したにもかかわらず、第 3.3.1.項に示したとおり、第 2 回学習後以降の自動評価スコアは毎回誤差レベルの微動にとどまっている。本例も大半の学習回で翻訳文が変化したが、全て一般用語や文構造のランダム的な変化によるものであり、突発的な不適訳語の採用を除けば、どれも翻訳品質を左右しない同義の範囲内のものであった。よって、自動評価の全体平均スコアから看取された翻訳品質の傾向は、少なくとも本例に関してはほぼ肯定される。

- ・ ただし、第 12 回学習後の全体平均スコアの顕著な向上が示す「みんなの自動翻訳」大規模アップデートの品質改善効果は、本例からは確認できなかった。本例の場合、アップデート以前から良好な品質で、改善の余地がなかったためと考えられる。

- ・ 一方、大規模アップデート以降の「不変」文の増加という傾向に関しては、本例においても大規模アップデート以降、一般用語や文構造のランダム的な変化が明らかに小規模・局所的になり、少数かつ同義の適訳語の間での変化に限定されるようになった。「みんなの自動翻訳」の訳語選定ロジックが、アップデートを機にランダム的な変化を抑制する方向に改良された可能性が高い。各文においてランダム的な訳語変化が抑制された結果、一部の文では翻訳文が変化しなくなり、「不変」文の増加をもたらしたと考えられる。本例に

関しても、第 16 回学習後及び第 20 回学習後は、前回と完全一致の翻訳文が出力された。

こうした変化により、本例では、大規模アップデート以前に度々の品質低下を招いていた「不適訳語の採用」が発生しなくなった。評価用コーパス全体で同様の改善がなされたことが、大規模アップデート以降の自動評価スコアの向上の一因である可能性が高い。

- ・ 自動評価の全体平均スコアが示した翻訳品質の推移は、本例の実際の品質変化とおおむね整合していた。ただし、本例の各回の機械翻訳文に付された文単位の自動評価スコアは実際の翻訳品質とかい離した変動を示すことが多く、個々の文単位の評価精度には限界があることもうかがえた。本例を見る限り、自動評価スコアはあくまで一定規模以上の母集合による平均スコアとして利用すべきものといえる。

こうした知見に関しては、後続の実例分析において引き続き検証し、補強・修正していく。

3.4.2. 実例 3-2：文番号#1110、汎用 NT／特許 NT

二つめの実例として取り上げる文番号#1110 は、無効宣告請求審決において、所定手続である「口頭審理」の経緯説明部分で多用される頻出文である。原文と正解訳文は以下のとおりである。

【文番号：#1110、汎用 NT／特許 NT】

原文	在口头审理过程中,
正解訳文	口頭審理の経緯は以下のとおりである：

本例は中国審決では頻出文であり、学習データ中には同一文が多数含まれている。本例において、こうした「学習データ中に同一文が存在する文」の挙動を確認する。下表に、汎用 NT、特許 NT それぞれの無学習時から第 20 回学習後まで全回の機械翻訳結果を一覧で示す。

	汎用 NT	特許 NT
#0	口頭審理の過程で	口頭監査プロセスにおいて
#1	口頭審理の経緯は以下のとおりである：	口頭審理の経緯は以下のとおりであった：
#2	口頭審理の内容は以下のとおりであった：	口頭審理の経緯は以下のとおりである：
#3	口頭審理の経緯は以下のとおりである：	口頭審理の内容は以下のとおりであった：
#4	口頭審理の経緯は以下のとおりである：	口頭審理において
#5	口頭審理の経緯は以下のとおりであった：	口頭審理の経緯は以下のとおりである：
#6	口頭審理の内容は以下のとおりであった：	口頭審理の内容は以下のとおりであった：
#7	口頭審理の経緯は以下のとおりである：	口頭審理の内容は以下のとおりであった：

#8	口頭審理において	口頭審理において
#9	口頭審理において	口頭審理において
#10	口頭審理の経緯は以下のとおりである:	口頭審理において
#11	口頭審理の経緯は以下のとおりである:	口頭審理の内容は以下のとおりである:
#12	口頭審理の経緯は以下のとおりである:	口頭審理の経緯は以下のとおりである:
#13	口頭審理の経緯は以下のとおりである:	口頭審理の内容は以下のとおりであった:
#14	口頭審理の経緯は以下のとおりである:	口頭審理の経緯は以下のとおりである:
#15	口頭審理の経緯は以下のとおりである:	口頭審理の経緯は以下のとおりである:
#16	口頭審理の経緯は以下のとおりである:	口頭審理の経緯は以下のとおりである:
#17	口頭審理の経緯は以下のとおりである:	口頭審理の経緯は以下のとおりである:
#18	口頭審理の経緯は以下のとおりである:	口頭審理の内容は以下のとおりであった:
#19	口頭審理の経緯は以下のとおりである:	口頭審理の経緯は以下のとおりである:
#20	口頭審理の経緯は以下のとおりである:	口頭審理の経緯は以下のとおりである:

無学習時の機械翻訳を見ると、専門性の高い専門用語である「口頭審理」が、前例では特許用語への対応度で劣っていた汎用 NT のほうで正解訳文どおり「口頭審理」と正しく訳されている。一方、特許 NT の訳語は「口頭監査プロセス」と不適訳である。ただし、これは汎用 NT が「口頭」「審理」をそれぞれ「口頭」「審理」と直訳した結果、偶然適訳語となったものと考えられる。前例でも無学習時の汎用 NT では特許・審決用語が直訳調で訳される傾向が顕著であった。

文構造に関しては、両エンジンとも無学習時は「…過程で」「…プロセスにおいて」と文が完結していない。これは、原文自体が「在口頭審理过程中、」と文の途中で強制改行された形になっているためであり、原文どおりといえる。

一方、冒頭に示した正解訳文、すなわち本事業で作成した人手翻訳文は、「口頭審理の経緯は以下のとおりである:」と終止形でひとまず文を結ぶ形に意識している。これは、この原文が本例のように強制改行されている場合、次行以降には口頭審理での複数の出来事が複数の文にわたり箇条書きされていることが多く²⁰、この形式で訳すほうが、この文が後続の箇条書き全体に係っていると理解しやすいためである。加えて、この文を「…過程で」「…プロセスにおいて」と直訳した場合、後続行の末尾で「…ということが明示された。」とい

²⁰ 本例もこれに該当する。具体的には評価用コーパス中の後続3文(文番号#1111~#1113)がこれにあたり、それぞれ(1)~(3)の付番を伴って箇条書きされている。仮に本例を無学習時の機械翻訳文「口頭審理の過程で」と直訳した場合、最終行(#1113)の末尾に「…ということが明示された。」というような縮めが必要となるが、原文にはこうした文言は存在せず、機械翻訳では事実上不可能である。

った締め文言を付加しないと文が完結しないが、後続行の原文にはそのような文言は存在しないため、機械翻訳でこれを達成することは難しい。

いずれにせよ、この正解訳文は、原文を直訳しただけでは得られない「意識文」である。これを踏まえて初回学習後の機械翻訳文を見ると、汎用 NT では初回学習後に早くも「**口頭審理の経緯は以下のとおりである:**」と正解訳文と完全一致の翻訳文が出力されている。この訳文は原文の直訳では到達できないものであるが、学習データに基づいて翻訳文を組み立てる「みんなの自動翻訳」では、意識であっても学習データ中で多用されていれば順当に採用されることが示された。一方、特許 NT の初回学習後の翻訳文も、正解訳文と完全一致ではないものの、「**口頭審理の経緯は以下のとおりであった:**」という同じ構造の意識文が採用されており、何ら遜色はない。

第 2 回学習後以降の各回の機械翻訳結果を概観すると、両エンジンとも「**口頭審理の経緯は以下のとおりである:**」という正解訳文どおりの基本文型を中心に、「**経緯**」と「**である**」がそれぞれ「**内容**」「**であった**」に置き換わった「**口頭審理の内容は以下のとおりであった:**」、「**である**」のみ「**であった**」に置き換わった「**口頭審理の経緯は以下のとおりであった:**」、「**経緯**」のみ「**内容**」に置き換わった「**口頭審理の内容は以下のとおりである:**」のいずれかが採用されていることが多い。つまり、前例（実例 3-1）で顕著であった「一般用語・表現が複数の候補間でランダム的に変化する」という傾向が、本例にも色濃く表れている。

また、両エンジンで採用される機械翻訳文のバリエーション自体は共通でも、それらが採用されるタイミングは同期しておらず、むしろ大規模アップデート以前の多くの回では、「**口頭審理の経緯は以下のとおりである:**」という基本文型と「**口頭審理の経緯は以下のとおりであった:**」等の変形パターンとが、両エンジンで逆方向に変化している。これも前例で見られた「学習データが同じであるにもかかわらず両エンジンで正反対の変化が生じる」という現象に相当する。

さらに大規模アップデート以前においては、無学習時の訳し方に近い「**口頭審理において**」という直訳調の翻訳文が、汎用 NT では第 8 回、第 9 回学習後、特許 NT では第 4 回、第 8 回、第 9 回、第 10 回学習後としばしば採用されている。中国審決コーパスデータは前述の理由でこの訳し方は採用しておらず²¹、学習データ中での出現頻度は低いはずだが、このように学習データの内容と直結していないような不適訳語が突発的に採用されるのも前例と同様である。

²¹ 厳密には「在口头审理过程中,」は直後で強制改行されずに後続文と連結されている場合もあり、こうしたケースでは「**口頭審理において、…**」と直訳調に訳されることも多い。

前例では、こうした一般用語や文構造のランダム的な変化や不適訳語の突発的な採用がニューラル翻訳方式における訳語選定のメカニズムに起因すること、このためあらゆる文において同様に発生する可能性が高いことを論じた。本例の状況も、この推論におおむね当てはまる。

また前例では、第 12 回時の大規模アップデート以降の「みんなの自動翻訳」は、おそらくロジックの改良によってこうした訳語のランダム的な変化が抑制され、訳語の的確性が向上していた。

この傾向の変化は本例にも等しく見られ、第 12 回学習後以降は、特許 NT の第 13 回後と第 18 回後というわずかな例外を除き、両エンジンほぼ全ての回で「**口頭審理の経緯は以下のとおりである:**」という基本文型が安定的に採用されるようになった。例外の 2 回も「口頭審理の**内容**は以下のとおり**であった:**」という、基本文型と同義・同構造の翻訳文であり、アップデート以前にしばしば採用されていた直訳調の「口頭審理**において**」などは一切採用されていない。つまり、前例における「機械翻訳文の変化が小規模・局所的になり、採用される訳語も適切かつ同義な少数の訳語候補に限定されるようになった」という傾向どおりであり、この傾向が各文に共通する全体的なものである可能性がさらに高くなった。なお、本例のこうした状況は第 3.3.2.項における大規模アップデート以降の「不変」文の顕著な増加とも整合する。

なお、本例中唯一の審決用語である「口頭審理」は、初回学習以降、両エンジンとも常に「口頭審理」という適訳語で完璧に安定した。この「特許・審決用語が学習初期から適訳語で常時安定する」傾向も前例と同じである。

自動評価スコアの妥当性に関しては、本例の場合、機械翻訳文が正解訳文と完全一致し、BLEU、RIBES とも満点 (1.000) が付された回が多かったが、それ以外の、例えば「口頭審理の**内容**は以下のとおり**であった:**」という機械翻訳文は、実質的な品質は同等であるにもかかわらず、BLEU スコアは 0.4367、RIBES は 0.9306 と、特に BLEU で非常に低いスコアとなっている。直訳調の「口頭審理**において**」も、採用すべき文構造ではないにせよ、BLEU で 0.0444 と極端に低いスコアが付されており、実際の品質と激しくかい離している。

以上、実例 3-2 の汎用 NT における各回の機械翻訳結果を定点観測した。最後に、実例 3-1 の観測結果から導出していた各傾向について、実例 3-2 の状況を照合する。

- ・ 特許・審決用語に関しては、本例では「口頭審理」一語のみが該当する。両エンジンとも、中国審決コーパスデータの初回学習後以降全ての回で「口頭審理」と適訳語が採用されており、特許・審決用語が学習初期から正解訳文どおりの適訳語で安定するという傾向は本例にも等しく当てはまった。
- ・ 一般用語や文構造に関しては、「口頭審理の経緯は以下のとおりである：」という基本文型をベースに「経緯」から「内容」への変化と、「である」から「であった」への変化という特定の箇所におけるランダム的な変化や、学習データの内容に準拠しない不適な文構造「口頭審理において」が散発的に採用されるなど、実例 3-1 と同様の傾向が見られた。
- ・ 第 12 回時の大規模アップデート以降は、基本文型「口頭審理の経緯は以下のとおりである：」が採用される確率が非常に高くなった。特許 NT の第 13 回、第 18 回学習後のみ「経緯」から「内容」への変化と「である」から「であった」への変化が発生したが、不適な文構造「口頭審理において」は一切採用されなくなり、一般用語や文構造のランダム的な変化が抑制されたことが見て取れた。
- ・ 本例では汎用 NT と特許 NT の双方のエンジンの全回の機械翻訳結果を示した。特に大規模アップデート以前は両エンジン共通のいくつかのパターンの翻訳文がランダム的に採用されており、回ごとの優劣はあるものの、全回を通して見ると両エンジン間の翻訳品質は同等とみなせる。この結果は、初回学習以降、両エンジンの自動評価の平均スコアが常時ほぼ同等であったことと整合しており、一定規模の母集合の平均を採れば自動評価スコアは十分な信頼性を有することが、前例に引き続き実証された。
- ・ ただしその一方で、本例に付された自動評価スコアは、実質的に同等の内容であるにもかかわらず正解訳文との表層的な一致度の違いにより不当かつ極端なスコアが付されることが多く（特に BLEU）、個々の文レベルの評価精度は高くないことが再確認された。

以上、前例で指摘した各種傾向は、実例 3-2 においてもほぼ該当した。

3.4.3. 実例 3-3 : 文番号#683、汎用 NT

ここまでに取り上げた 2 件の実例には技術用語が含まれていなかったため、実例 3-3 には技術用語を含む文番号#683 を取り上げる。汎用 NT における全 21 回の機械翻訳結果を一覧で示す。

【文番号：#683、汎用 NT】

原文	④对于乳液聚合而言, 需要破乳步骤, 悬浮聚合没有破乳步骤, 而对比文件 4 完全没有记载破乳步骤。	
正解訳文	④乳化重合については、解乳化ステップが必要であり、懸濁重合には解乳化ステップはなく、また引用文献 4 には解乳化ステップの記載は全くない。	
無学習	④乳化重合に対して、乳化破壊ステップが必要であり、懸濁重合は乳化ステップがなく、対照文書 4 は完全に乳化破壊プロセスを記載しない。	
	—	BLEU : 0.1829 (—) RIBES : 0.7480 (—)
第 1 回学習後	④乳化重合については、乳化を防止するステップを必要とし、懸濁重合には乳化ステップがないが、引用文献 4 には全く記載されていない。	
	悪化	BLEU : 0.3216 (+0.1387) RIBES : 0.7926 (+0.0446)
第 2 回学習後	④乳化重合については、乳化重合ステップが必要であり、懸濁重合は乳化ステップがないため、引用文献 4 には乳化破壊ステップが全く記載されていない。	
	同等	BLEU : 0.3991 (+0.0775) RIBES : 0.8177 (+0.0251)
第 3 回学習後	④エマルジョン重合については、乳化重合ステップが必要とされ、懸濁重合はないが、引用文献 4 には解乳化ステップが全く記載されていない。	
	同等	BLEU : 0.3289 (-0.0702) RIBES : 0.7982 (-0.0195)
第 4 回学習後	④乳化重合については、乳化重合ステップを必要とし、懸濁重合は乳化ステップを行わず、引用文献 4 には乳化破壊ステップが全く記載されていない。	
	同等	BLEU : 0.2567 (-0.0722) RIBES : 0.7434 (-0.0548)
第 5 回学習後	④乳化重合については、解乳化ステップを必要とし、懸濁重合は乳化ステップを有さず、引用文献 4 には解乳化のステップが全く記載されていない。	
	同等	BLEU : 0.3626 (+0.1059) RIBES : 0.8290 (+0.0856)
第 6 回学習後	④エマルジョン重合については、乳化ステップを必要とし、懸濁重合は乳化ステップがなく、引用文献 4 には乳化ステップが全く記載されていない。	

	悪化	BLEU : 0.2342 (-0.1284) RIBES : 0.8294 (-0.0004)
第 7 回学習後		④エマルション重合について,乳化ステップを必要とし,懸濁重合には乳化ステップが記載されていないが,引用文献 4 は完全に乳化ステップを記載していない。
	同等	BLEU : 0.2149 (-0.0193) RIBES : 0.6940 (-0.1354)
第 8 回学習後		④エマルション重合について,乳化ステップは必要であり,懸濁重合は乳化ステップがなく,引用文献 4 には乳化ステップが全く記載されていない。
	同等	BLEU : 0.3422 (+0.1273) RIBES : 0.8170 (+0.1230)
第 9 回学習後		④エマルション重合については,乳化ステップを必要とし,懸濁重合は乳化ステップを有さないが,引用文献 4 には乳化ステップが全く記載されていない。
	同等	BLEU : 0.2280 (-0.1142) RIBES : 0.7019 (-0.1151)
第 10 回学習後		④エマルション重合の場合,乳化ステップが必要であり,懸濁重合には乳化ステップがなく,引用文献 4 には粉乳ステップが完全に記載されていない。
	同等	BLEU : 0.4189 (+0.1909) RIBES : 0.8175 (+0.1156)
第 11 回学習後		④エマルション重合について,乳化ステップを必要とし,懸濁重合に解乳化ステップがなく,一方、引用文献 4 には解乳化ステップが全く記載されていない。
	良化	BLEU : 0.3739 (-0.0450) RIBES : 0.8620 (+0.0445)
第 12 回学習後		④乳化重合について,解乳化ステップが必要であり,懸濁重合には解乳化ステップがなく,引用文献 4 には解乳化ステップが全く記載されていない。
	良化	BLEU : 0.6563 (+0.2824) RIBES : 0.9596 (+0.0976)
第 13 回学習後		④乳化重合について,解乳化ステップが必要であり,懸濁重合には解乳化ステップがなく,引用文献 4 には解乳化ステップが全く記載されていない。
	同等	BLEU : 0.6563 (±0) RIBES : 0.9596 (±0)
第 14 回学習後		④乳化重合については,解乳化ステップが必要であり,懸濁重合には解乳化ステップがないが,引用文献 4 には解乳化ステップが全く記載されていない。
	悪化	BLEU : 0.6190 (-0.0373) RIBES : 0.9342 (-0.0254)
第 15 回学習後		④乳化重合については,解乳化ステップが必要であり,懸濁重合には解乳化ステップがないが,引用文献 4 には解乳化ステップが全く記載されていない。

	同等	BLEU : 0.6190 (±0) RIBES : 0.9342 (±0)
第 16 回学習後	④乳化重合については、解乳化ステップが必要であり、懸濁重合には解乳化ステップがないが、引用文献 4 には解乳化ステップが全く記載されていない。	
	同等	BLEU : 0.6190 (±0) RIBES : 0.9342 (±0)
第 17 回学習後	④乳化重合について、解乳化ステップを必要とし、懸濁重合には解乳化ステップがないが、引用文献 4 には解乳化ステップが全く記載されていない。	
	同等	BLEU : 0.5126 (-0.1064) RIBES : 0.9172 (-0.0170)
第 18 回学習後	④乳化重合について、解乳化ステップを必要とし、懸濁重合には解乳化ステップがないが、引用文献 4 には解乳化ステップが全く記載されていない。	
	同等	BLEU : 0.5126 (±0) RIBES : 0.9172 (±0)
第 19 回学習後	④乳化重合については、解乳化ステップが必要であり、懸濁重合には解乳化ステップがなく、引用文献 4 には解乳化ステップが全く記載されていない。	
	同等	BLEU : 0.6446 (+0.1320) RIBES : 0.9538 (+0.0366)
第 20 回学習後	④乳化重合について、解乳化ステップが必要であり、懸濁重合には解乳化ステップがなく、引用文献 4 には解乳化ステップが全く記載されていない。	
	同等	BLEU : 0.6563 (+0.0117) RIBES : 0.9596 (+0.0058)

本例は、技術用語として「乳液聚合（乳化重合）」、「破乳步骤（解乳化ステップ）」「悬浮聚合（懸濁重合）」の 3 語を含む（「破乳步骤」は文中に三回出現）。これら 3 語のうち「悬浮聚合」は、無学習時から最終学習後まで「懸濁重合」と正解訳文どおりの適訳語で常時安定していた。

一方、「乳液聚合」は、無学習時から第 5 回学習後までは正解訳文どおりに「乳化重合」と訳されていたが、第 6 回後から第 11 回後にかけては「エマルジョン重合」に変化し、第 12 回後以降は再び「乳化重合」が採用されるようになった。これら二種の訳語は、特許公報での使用頻度は「乳化重合」のほうが高いが、同義語であって内容理解上の優劣はない。

一方、原文中に三度出現する「破乳步骤」に関しては、特に学習初期には学習前後のみならず同一文中でも訳語変化が激しく、かつ誤訳も多発している。以下、各回における「破乳步骤」の訳語の変遷を一覧で示す。表では、原文での出現順に①～③と付番している。

表 3.4.1. 実例#683 における技術用語「破乳步骤」の訳語変化状況（汎用 NT）

原語	破乳步骤①	破乳步骤②	破乳步骤③
正解訳	解乳化ステップ	解乳化ステップ	解乳化ステップ
無学習	乳化破壊ステップ	乳化ステップ	乳化破壊プロセス
第 1 回	乳化を防止するステップ	乳化ステップ	(欠落)
第 2 回	乳化重合ステップ	乳化ステップ	乳化破壊ステップ
第 3 回	乳化重合ステップ。	(欠落)	解乳化ステップ
第 4 回	乳化重合ステップ	乳化ステップ	乳化破壊ステップ
第 5 回	解乳化ステップ	乳化ステップ	解乳化のステップ
第 6 回	乳化ステップ	乳化ステップ	乳化ステップ
第 7 回	乳化ステップ	乳化ステップ	乳化ステップ
第 8 回	乳化ステップ	乳化ステップ	乳化ステップ
第 9 回	乳化ステップ	乳化ステップ	乳化ステップ
第 10 回	乳化ステップ	乳化ステップ	粉乳ステップ
第 11 回	乳化ステップ	解乳化ステップ	解乳化ステップ
第 12 回	解乳化ステップ	解乳化ステップ	解乳化ステップ
第 13 回	解乳化ステップ	解乳化ステップ	解乳化ステップ
第 14 回	解乳化ステップ	解乳化ステップ	解乳化ステップ
第 15 回	解乳化ステップ	解乳化ステップ	解乳化ステップ
第 16 回	解乳化ステップ	解乳化ステップ	解乳化ステップ
第 17 回	解乳化ステップ	解乳化ステップ	解乳化ステップ
第 18 回	解乳化ステップ	解乳化ステップ	解乳化ステップ
第 19 回	解乳化ステップ	解乳化ステップ	解乳化ステップ
第 20 回	解乳化ステップ	解乳化ステップ	解乳化ステップ

各学習回の機械翻訳文における「破乳步骤」の訳語の変遷を見ていくと、まず無学習時の翻訳文では、①が「乳化破壊ステップ」、②が「乳化ステップ」、③が「乳化破壊プロセス」とそれぞれ異なる訳語が採用されている。このうち①「乳化破壊ステップ」と③「乳化破壊プロセス」は、同じ技術用語が一文中で訳ゆれしているという問題を除けばほぼ同義であり、正解訳文の「解乳化ステップ」と同等の理解に到達可能であるが、②の「乳化ステップ」は意味が正反対であり、重大な誤訳である。

続く初回学習後の翻訳文では、①の訳語が「乳化を防止するステップ」に変化し、②は「乳化ステップ」と引き続き誤訳、そして③は翻訳文から欠落してしまっている。無学習時は①、③とも、正解訳文とは異なるが正しい意味を読み取れる訳語が採用されていたが、今回の①

の訳語「**乳化を防止するステップ**」は意味が大きく逸れており、改悪である。③についても翻訳文から完全に欠落しており、状況は無学習時よりも悪化している。本調査分析では、それまで訳出されていた語がこのように学習後に突発的に欠落するケースは他の文でも散見された。

続く第2回学習後の翻訳文では、「**乳化重合ステップ**」「**乳化ステップ**」「**乳化破壊ステップ**」と再び①～③でそれぞれ別の訳語が採用された。②の「**乳化ステップ**」の誤訳は今回も変わらなかったが、③では前回発生していた欠落が解消し、初回学習時に①で採用されていた「**乳化破壊ステップ**」が採用された。このように、もともと訳出されていた語が突発的に欠落したような場合、次回学習後には自然に解消されることが多い。

一方、①では今回「**乳化重合ステップ**」という初出の訳語が採用された。もちろん誤訳であるが、原語である「**破乳ステップ**」には「**重合**」と誤訳されるような要素は存在せず、例えば前回の「**乳化を防止するステップ**」のような通常の誤訳とは性質が異なっている。

原文では、「**重合**」に相当する「**聚合**」という語が「**乳液聚合**」「**懸浮聚合**」として2回出現する。ただし、これらは第2回時の翻訳文でも「**乳化重合**」「**懸濁重合**」と正しい位置で正しく訳出されており、これら2つの「**聚合**」のどちらかに相当する「**重合**」が構文解析の誤りで「**破乳ステップ**」の位置に出力されたわけではない。つまり、①の訳語「**乳化重合ステップ**」に含まれる3つめの「**重合**」は、原文中には存在しない語句と見なされる。原文中に存在しない語句が翻訳文中に混入することを「湧き出しエラー」と呼ぶが、「みんなの自動翻訳」ではこのように、何らかの理由で翻訳文から本来あるべき語句が欠落してしまった場合、埋め合わせとしてその位置に文中の別の語句を重複出力するタイプの湧き出しエラーが散見される。この湧き出しエラーに関しては、第4.5.5.7.項で詳しく分析する。

続く第3回学習後の翻訳文では、①は前回に引き続き「**乳化重合ステップ**」と湧き出しエラーを含む訳語が出力され、一方、これまで「**乳化ステップ**」と誤訳されてきた②は今回、突発的に欠落している。そして③では、正解訳文どおりの「**解乳化ステップ**」という訳語が初めて採用された。第3回時までに学習させた中国審決コーパスデータ中には「**破乳ステップ：解乳化ステップ**」という語を含む文対は存在しないが、「**破乳**」を「**解乳化**」と訳している文対は第2回時に学習させたコーパス中に6文存在しており（全て審決番号32050 復審由来）、これらによって「**解乳化**」という訳語が新規に学習された可能性がある²²。学習された

²² デフォルト学習データの内容は非公開であるため、そこに「破乳」を「解乳化」と訳している事例が含まれていないとは断定できない。訳語候補としては当初から存在していたが、何らかの理由で第3回時

訳語がその回では採用されず第 3 回になってから採用された理由は判然としないが、各訳語に付されるベクトルは学習データ追加の都度再計算されるため、何かしら「解乳化」という訳語の採用を促す変化が生じたものと推察される。

なお、第 3 回以降は「破乳」を含むコーパスは第 10 回時まで存在せず、したがって学習データ中の「破乳」の出現頻度は第 2 回時から第 9 回時まで変化していない²³。それにもかかわらず、第 3 回学習後に採用された「解乳化ステップ」は直後の第 4 回後には採用されず、第 5 回後には再度①と③で採用されたものの、その後は第 11 回後に②と③で採用されるまで長らく不採用となっている。単純に中国審決コーパスデータ上の出現頻度に基づけば「破乳」の訳語には常に「解乳化」が採用されるはずであり、この状況は、「みんなの自動翻訳」における訳語選定の基準、すなわち学習データ追加の都度再計算されるベクトルの値が、必ずしもコーパス中の出現頻度に直結したものではないことを示している。

ただし、このような学習データの内容に直結していないと映る訳語選定の傾向は、「みんなの自動翻訳」の大規模アップデートが実施された第 12 回学習時以降、本例においては明らかに変化した。アップデート直後の第 12 回学習後では初めて①～③の訳語が全て「解乳化ステップ」という学習データに準拠した適訳語で統一され、それ以降、全ての回で「破乳手順」の訳語は①～③とも「解乳化ステップ」で完全に安定した。

調査したところ、第 12 回時に追加したコーパス中で「破乳」を「解乳化」と訳している文対は 1 件のみであり、全学習データ中でこの訳例の出現頻度が劇的に変化したわけではない。同種の訳例は第 2 回学習の時点から学習データ中に複数件存在していたことを考えると、むしろ大規模アップデート以前は、こうした学習データの内容が訳語選定に的確に反映できていなかった可能性が高い。つまり、学習データから「破乳」の訳語として「解乳化」という訳語候補は得られてはいたが、それが安定して採用されるような訳語選定のロジックとなっていなかったものと推察される。

そして本例の第 12 回時以降の翻訳文の状況からは、大規模アップデート時に訳語選定のロジックが改良され、学習データの内容がよりの確に訳語選定に反映されるようになったことで、学習データ中で最も有力な訳語候補が毎回順当に採用されるようになったものと考えられる。つまりエンジンのロジックの改良、具体的には各訳語に付されるベクトルの精

までは不採用であったという可能性もある。

²³ ただし「みんなの自動翻訳」の定期バージョンアップ時の学習データの追加などで状況が変化している可能性は存在する。

度向上により、同じ学習データから、より適切な翻訳文を生成できるようになったということである。先に分析した実例 3-1～3-2 の状況も、この推論と矛盾していない。

以上、本例は技術用語の訳語変化に焦点を絞って分析した。最後に、本例から得た技術用語に関する新たな知見と、前二例から把握された各種傾向を本例と照合した検証結果についてまとめる。

- ・ 本例における技術用語は、「**懸濁重合**」のように無学習時から訳語が安定しているもの、「**乳液重合**」のように（前二例の一般用語と同様）同義の範囲内でランダム的に変化するもの、そして「**破乳ステップ**」のように各学習回、そして同一回の同一文中でも訳語が不安定に変化し、その多くが誤訳・不適訳となっているものと三者三様であった。
- ・ 「**破乳ステップ**」に関しては、学習データ中には早期から適訳事例が存在していたにもかかわらず、大規模アップデート以前の「みんなの自動翻訳」ではその適訳語が採用されることはまれであり、学習データの内容が必ずしも直接かつ的確に訳語選定に反映されていない状況がうかがえた。
- ・ これに対し、大規模アップデートを経た第 12 回学習後以降は、「**破乳ステップ**」は学習データに準拠した「**解乳化ステップ**」という訳語で完璧に安定するようになった。第 12 回時の学習データ自体には「**破乳ステップ**」という用語に関する劇的な変化はなく、この傾向の変化は「みんなの自動翻訳」の訳語選定ロジック、多次元ベクトルの再計算ロジック自体が大規模アップデートを機に改良され、学習データの内容がよりの確に訳語選定に反映されるようになったことによるものと推察される。この推論は実例 3-1～3-2 の状況とも合致する。
- ・ 特許・審決用語に関しては、本例では「対比文書 4」一語のみが該当する。無学習時の汎用 NT ではこれを「対照文書 4」と誤訳していたが、中国審決コーパスデータの初回学習以降は適訳語である「引用文献 4」が安定して採用され続けた。
- ・ 一方、一般用語や文構造に関しては、前二例と同様、同義の範囲内でのランダム的な変化が見られた。大規模アップデート以降にこうした変化が抑制された点も同様であり、第 12 回学習後以降は、各回の翻訳文の変化箇所は 2 箇所のみ、具体的には「**乳液重合について**」と「**乳液重合については**」との間の変化、及び「**解乳化ステップが必要であり**」と「**解乳化ステップを必要とし**」との間の変化のみに局所化・小規模化された。いずれも

完全に同義の範囲での変化であり、したがって大規模アップデート以降の各回の翻訳品質は完全に同等であった。

- ・ 本例では、技術用語「破乳步骤」が適訳語「解乳化ステップ」で安定したことにより、大規模アップデート以降は実質的な翻訳品質が顕著に改善され、最終学習時まで常時その品質が維持された。この結果は、自動評価の全体平均スコアが示す翻訳品質の推移と合致する。

本例における大規模アップデート以降の品質改善は、学習データの内容がよりの確に訳語選定に反映されるようになったことに起因している。換言すれば訳語選定におけるランダム性が減じ、突発的な不適訳語の採用が減少したためである。前二例（実例 3-1～3-2）では大規模アップデート前後で特段の品質変化は見られなかったが、両例ともアップデート以降の訳語選定のランダム性の抑制は確認されており、機械翻訳に見られる傾向自体は各実例で共通している。

こうした状況から、大規模アップデート以降の全体平均スコアの顕著な向上は、本例のようにランダム的な訳語選定が品質上のネックとなっていた多くの文において、適訳語が安定して採用されるようになったことで翻訳品質が底上げされ、その後も常時維持された結果と考えられる。

3.4.4. 実例 3-4 : 文番号#1995、汎用 NT

続いての実例は、特許・審決用語を全く含まない技術内容文、具体的には、特許請求の範囲を摘記した箇所からの一文である。以下、各回の機械翻訳文を示す。

【文番号：#1995、汎用 NT】

原文	向学员提供与所述第一驾驶学习课件的内容相匹配的第一驾驶模拟练习。”	
正解訳文	学生に前記第一運転学習コースウェアの内容に見合う第一運転模擬練習を提供する。」	
無学習	第一運転学習コースウェアの内容にマッチした第一運転模擬練習を学習者に提供する。」	
	—	BLEU : 0.5894 (—) RIBES : 0.9306 (—)
第 1 回学習後	前記第一運転学習コースウェアの内容に 一致する 第一運転 シミュレーション演習 を学習者に提供する。」	
	同等	BLEU : 0.5034 (-0.0860) RIBES : 0.8739 (-0.0567)
第 2 回学習後	学習者に 前記第一運転学習コースウェアの内容に一致する第一運転 シミュレーション を提供する。」	
	同等	BLEU : 0.6086 (+0.1052) RIBES : 0.8987 (+0.0248)
第 3 回学習後	学習者に前記第一運転 学習授業 の内容に マッチングする 第一運転 模擬演習 を提供する。」	
	劣化	BLEU : 0.4934 (-0.1152) RIBES : 0.9306 (+0.0319)
第 4 回学習後	前記第一運転 学習教材 の内容にマッチングする第一運転 シミュレーション演習 を 学習者に 提供する。」	
	同等	BLEU : 0.3523 (-0.1411) RIBES : 0.8579 (-0.0727)
第 5 回学習後	学習者に 前記第一運転 学習ファイル の内容に一致する第一運転 シミュレーション学習 を提供する。」	
	同等	BLEU : 0.4444 (+0.0921) RIBES : 0.8739 (+0.0160)
第 6 回学習後	前記第一運転 学習部 の内容に 適合する 第一運転 シミュレーション演習 を オペレータ に提供する。」	
	劣化	BLEU : 0.3687 (-0.0757) RIBES : 0.8671 (-0.0068)
第 7 回学習後	学習者に 前記第一運転 学習ウィジェット の内容と一致する第一運転 シミュレーション学習 を提供する。」	
	向上	BLEU : 0.4166 (+0.0479) RIBES : 0.8579 (-0.0092)
第 8 回学習後	学習者に前記第 1 運転 学習学習手段 の コンテンツ に マッチングする 第 1 運転 シミュレーション学習 を提供する。」	

	劣化	BLEU : 0.2254 (-0.1912) RIBES : 0.8144 (-0.0435)
第 9 回学習後	学習者に前記第一運転学習学習部品の内容に適合する第一運転シミュレーション演習を提供する。」	
	向上	BLEU : 0.4254 (+0.2000) RIBES : 0.8651 (+0.0537)
第 10 回学習後	学習者に前記第一運転学習教材の内容にマッチングする第一運転シミュレーション演習を提供する。」	
	向上	BLEU : 0.4444 (+0.0190) RIBES : 0.8739 (+0.0088)
第 11 回学習後	学習者に前記第一運転学習教材の内容とマッチングする第一運転シミュレーション演習を提供する。」	
	同等	BLEU : 0.4166 (-0.0278) RIBES : 0.8579 (-0.0160)
第 12 回学習後	前記第一運転学習教材の内容に適合する第一運転シミュレーション演習を学習者に提供する。」	
	向上	BLEU : 0.3523 (-0.0643) RIBES : 0.8579 (±0)
第 13 回学習後	学習者に前記第一運転学習教材の内容第一運転シミュレーション練習を提供する。」	
	劣化	BLEU : 0.5119 (+0.1596) RIBES : 0.8993 (+0.0414)
第 14 回学習後	訓練生に前記第一運転学習教材の内容とマッチングする第一運転シミュレーション練習を提供する。」	
	向上	BLEU : 0.4651 (-0.0468) RIBES : 0.8739 (-0.0254)
第 15 回学習後	訓練生に前記第一運転学習ウィジェットの内容にマッチングする第一運転シミュレーション練習を提供する。」	
	同等	BLEU : 0.4934 (+0.0283) RIBES : 0.8891 (+0.0152)
第 16 回学習後	訓練生に前記第一運転学習教材の内容に一致する第一運転シミュレーション練習を提供する。」	
	向上	BLEU : 0.4934 (±0) RIBES : 0.8891 (±0)
第 17 回学習後	訓練生に前記第一運転学習教材の内容とマッチングする第一運転シミュレーション練習を提供する。」	
	劣化	BLEU : 0.4651 (-0.0283) RIBES : 0.8739 (-0.0152)
第 18 回学習後	前記第一運転学習教材の内容と一致する第一運転シミュレーション練習を訓練生に提供する。」	
	向上	BLEU : 0.3389 (-0.1262) RIBES : 0.8579 (-0.0160)
第 19 回学習後	訓練生に前記第一運転学習教材の内容にマッチングする第一運転シミュレーション練習を提供する。」	
	劣化	BLEU : 0.5002 (+0.1613) RIBES : 0.9306 (+0.0727)

第 20 回学習後	前記第一運転学習教材の内容に適合する第一運転シミュレーション練習を訓練生に提供する。」	
	向上	BLEU : 0.3785 (+0.1217) RIBES : 0.9036 (-0.0270)

本例には「学员 (学生)」「学習课件 (学習コースウェア)」「匹配的 (見合う)」「模擬練習 (模擬練習)」と 4 語の技術用語が含まれているが、いずれの訳語も、前例における「破乳 歩驟」の訳語と同様、学習前後で、複数の訳語候補間で不安定な変化を繰り返している。以下、各語の各回学習後での訳語を一覧表にまとめた。

表 3.4.2. 実例#1995 における技術用語の訳語変化状況 (汎用 NT)

原語	学员	学習课件	匹配的	模擬練習
正解訳	学生	学習コースウェア	見合う	模擬練習
無学習	学習者	学習コースウェア	マッチした	模擬練習
第 1 回	学習者	学習コースウェア	一致する	シミュレーション演習
第 2 回	学習者	学習コースウェア	一致する	シミュレーション
第 3 回	学習者	学習授業	マッチングする	模擬演習
第 4 回	学習者	学習教材	マッチングする	シミュレーション演習
第 5 回	学習者	学習ファイル	一致する	シミュレーション学習
第 6 回	オペレータ	学習部	適合する	シミュレーション演習
第 7 回	学習者	学習ウィジェット	一致する	シミュレーション学習
第 8 回	学習者	学習学習手段	マッチングする	シミュレーション学習
第 9 回	学習者	学習学習部品	適合する	シミュレーション演習
第 10 回	学習者	学習教材	マッチングする	シミュレーション演習
第 11 回	学習者	学習教材	マッチングする	シミュレーション演習
第 12 回	学習者	学習教材	適合する	シミュレーション演習
第 13 回	学習者	学習教材	(欠落)	シミュレーション練習
第 14 回	訓練生	学習教材	マッチングする	シミュレーション練習
第 15 回	訓練生	学習ウィジェット	マッチングする	シミュレーション練習
第 16 回	訓練生	学習教材	一致する	シミュレーション練習
第 17 回	訓練生	学習教材	マッチングする	シミュレーション練習
第 18 回	訓練生	学習教材	一致する	シミュレーション練習
第 19 回	訓練生	学習教材	マッチングする	シミュレーション練習
第 20 回	訓練生	学習教材	適合する	シミュレーション練習

「**学员** (正解訳文では「**学生**」。以下同じ)」に関しては、第 13 回学習後までは「**学習者**」、第 14 回後以降は「**訓練生**」と、学習前期と後期それぞれで、正解訳文と同義の訳語でほぼ安定している。ただし第 6 回後のみは「**オペレータ**」という不適訳語が突発的に採用された。

各学習回の学習データにおける「**学员**」の出現頻度を調査したところ、第 10 回学習時までの中国審決コーパスデータには「**学员**」という語自体が含まれていなかった。よって、無学習時から採用されていた「**学習者**」だけでなく、第 6 回後に採用された「**オペレータ**」も「みんなの自動翻訳」のデフォルト学習データに由来していると思わせる。つまり第 6 回時は、デフォルト学習データ中に訳語候補として存在していたがそれまで全く採用されなかったマイナーな訳語が 1 回のみ突発的に採用されたことになる。このような訳語選定のランダム性は大規模アップデート以前の「みんなの自動翻訳」には頻繁に見られ、例えば前例 3-3 でも「**破乳步骤 (③)**」の訳語として第 10 回後のみ「**粉乳ステップ**」というマイナーな訳語が唐突に採用されている。

その後の学習データでは、第 11 回学習時に「**学员**」を「**メンバー**」と訳した訳例が 1 文、第 13 回学習時には「**訓練生**」と訳した訳例が 11 文存在した。前者は訳語選定には反映されなかったが、後者は第 14 回後以降の「**学習者**」から「**訓練生**」への訳語変化の原因となったと考えられる。この結果は、「大規模アップデート以降、学習データ中の有力な訳語が順当に採用される確率が高くなった」という前例での考察と整合する。ただし、「**訓練生**」が採用されるようになったのが第 13 回当時ではなく、1 回遅れの第 14 回学習後からであった理由は不明である。

2 つめの技術用語「**学习课件 (学習コースウェア)**」は、中国審決コーパスデータ中には訳例は存在しない。ただし各回の訳語を見ると、「**学习**」は常時「**学習**」と訳されており、訳語が変化しているのは「**课件**」の部分のみである。この「**课件**」単独であれば、第 17 回時の学習データに「**教育ソフト**」と訳した事例が 2 文含まれていた。しかしこの訳語は翻訳文には採用されておらず、「**学员：メンバー**」と同様、文例が 1~2 文と少数であったため有力な訳語候補として扱われなかったものと推察される。

この「**学习课件**」は、大規模アップデート以前の第 11 回学習後までに「**课件**」部分の訳語が全 8 種類²⁴と多様に変化しており、その多くは 1 回のみ採用であるなど、「**学员**」よりもランダム的な訳語変化が顕著であった。この期間の中国審決コーパスデータ中に「**学习**

²⁴ 「コースウェア (#0~#2)」「授業 (#3)」「教材 (#4、#10~#11)」「ファイル (#5)」「部 (#6)」「ウィジェット (#7)」「学習手段 (#8)」「学習部品 (#9)」。

「**课件**」や「**课件**」の訳例は存在しておらず、全ての訳語がデフォルトの学習データから学習されたものと見られる。デフォルト学習データは常時ほぼ不変であり、かつ中国審決コーパスデータ中にも直接関係する訳例が存在しないにもかかわらず、学習の都度訳語が変化している状況から、やはり大規模アップデート以前の「みんなの自動翻訳」における訳語選定ロジック、すなわち各訳語のベクトル算出ロジックは、学習データ中の出現頻度と直結していなかったものと推察される。

本語に関しては、大規模アップデートが実施された第 12 回以降は傾向が変化し、第 15 回学習後の「学習ウィジェット」を除き、他の全ての回で「学習教材」という訳語が安定して採用されるようになった。ここまでの各例で考察したとおり、大規模アップデートによって「みんなの自動翻訳」の訳語選定ロジックが改良され、ランダム的な訳語変化が抑制されることで、学習データ中で最も有力な訳語候補が順当に採用される確率が高められたことが本例からもわかる。

3 つめの技術用語「**匹配的 (見合う)**」に関しては、当初から「適合する」「マッチングする」「一致する」という 3 種の訳語候補がランダム的に採用されていたが、アップデート以前とこの傾向はさほど変わらなかった。大規模アップデート時のロジックの改良によって訳語の安定性が向上したとはいえ、用語によっては、本語のように引き続き複数の訳語候補間で変化を続けるものも存在する。ここまで見てきた各実例でも、最終学習後まで同義の訳語間で不安定な変化を続けた語も局所的には存在していた。実際、大規模アップデート以降も全体の 3 分の 2 の文は学習の都度何らかの変化を続けており (⇒3.3.2.)、学習データ中に有力な訳語候補が複数存在するような用語に関しては、本例の「**匹配的**」と同様、大規模アップデート以降も訳語のランダム的な変化が継続していると考えべきである。第 3.4.1. 項で論じたニューラル機械翻訳方式の訳語選定の方法論からも、このように考えるのが妥当である。

最後の「**模拟练习 (模擬練習)**」に関しても、この単位では中国審決コーパスデータ中に訳例は存在しないが、「**模拟**」「**练习**」に分解すればどちらも多数の訳例が初回学習後の時点から存在していた。各学習回の訳語の変遷を見ていくと、「**模拟**」は学習初期を除けば「シミュレーション」が常時採用されており、一方「**练习**」も、大規模アップデート以前は「**演習**」と「**学習**」とがランダムに入れ替わっていたが、アップデート以降は第 12 回後を除き「**練習**」が安定して採用されるようになった。その結果、第 13 回後以降の訳語は毎回「シミュレーション練習」となっている。

中国審決コーパスデータ中の「**练习**」の訳語の出現頻度を見ると、全 86 文中 64 文が「**練習**」であり、学習初期に採用されていた「**演習**」と「**学習**」はそれぞれ 1 文、3 文と少数で

あった。つまり中国審決コーパスデータにおいては「**练习**」の訳語は「練習」の出現頻度が圧倒的に高く、第13回学習後以降の「練習」への訳語変化も、大規模アップデート時のロジックの改善によって、学習データ中の最有力な訳語が順当に採用されるようになった結果であると考えられる。

以上が、文番号#1995の汎用NTにおける訳語変化の状況であるが、ここまで論じてきた考察に基づけば、同様の状況が特許NTでも等しく見られるはずである。これを検証するため、特許NTに関しても同じ4語の訳語の変遷を一覧表にまとめた。

表 3.4.3. 実例#1995 における技術用語の訳語変化状況 (特許 NT)

原語	学员	学習课件	匹配的	模拟练习
正解訳	学生	学習コースウェア	見合う	模擬練習
無学習	生徒	学習コースウェア	一致する	シミュレーション練習
第1回	生徒	学習コース材	マッチする	シミュレーション
第2回	運転者	学習部材	一致する	シミュレーション
第3回	生徒	学習部材	適合する	アナログトレーニング
第4回	事業者	学習授業	一致する	シミュレーション
第5回	教学者	学習者	一致する	シミュレーション
第6回	生徒	学習部材	マッチングする	シミュレーション
第7回	学習員	学習ウィジェット	一致する	シミュレーション練習
第8回	学習者	学習コース	適合する	シミュレーショントレーニング
第9回	学習者	学習教材	マッチングする	シミュレーショントレーニング
第10回	学習者	学習教材	適合する	シミュレーション
第11回	学習者	学習ウィジェット	マッチングする	シミュレーション練習
第12回	学生	学習教材	マッチングする	シミュレーション練習
第13回	生徒	学習教材	マッチングする	シミュレーション練習
第14回	訓練生	学習教材	一致する	シミュレーション練習
第15回	訓練生	学習教材	一致する	シミュレーション練習
第16回	訓練生	学習教材	一致する	シミュレーション練習
第17回	訓練生	学習教材	一致する	シミュレーション練習
第18回	訓練生	学習教材	マッチングする	シミュレーション練習
第19回	訓練生	学習 記述	適合する	模擬練習
第20回	訓練生	学習教材	一致する	模擬 エクササイズ

表中のマーカー表示は、汎用 NT と同一の訳語が採用されていることを示している。一見して、大規模アップデート以前は各語とも両エンジンで訳語がほとんど一致しておらず、当時の「みんなの自動翻訳」における訳語のランダム性が顕著に表れている。特許 NT では、例えば「学员」の訳語として「**運転者**」「**事業者**」「**教学者**」など汎用 NT では採用されていなかった不適訳語が散発的に採用されており、両エンジンの基礎学習データの違いが影響している部分もあるであろうが、状況としては汎用 NT で第 6 回学習後のみ「**オペレータ**」が採用されたのと類似しており、学習データ中のマイナーな訳語候補がランダム的な訳語選定によって突発的に採用されることがあるという傾向が両者共通であることを示している。

大規模アップデートが実施された第 12 回以降は、「学员」「学習课件」「模拟练习」の 3 語はそれぞれ汎用 NT と同じ「訓練生」「学習教材」「シミュレーション練習」という最有力訳語が比較的安定して採用されるようになった。両エンジンとも（基礎学習データの一部を除き）同じ学習データを使用しているので、学習データに忠実な訳語選定が行われるようにロジックが改良されたのであれば、このように両エンジンで同じ訳語が採用されるようになるケースが増加するのが自然である。この結果は、本項で述べてきた考察内容の妥当性を裏付けるものといえるであろう。さらには、大規模アップデートにより訳語選定のランダム性が抑制されたことで、各エンジンの学習前後の訳語の安定化に加え、両エンジン間における訳語の均質化も進んだことを示す結果でもある。

ただし、「匹配的」の 1 語のみは、「一致する」「マッチングする」という訳語候補自体は汎用 NT と共通するが、この 2 種の訳語候補がほぼ毎回、両エンジン間で食い違っている。本語に関しては、汎用 NT の考察において、学習データ中に有力な訳語候補が複数存在するような用語に関してはアップデート以降も「みんなの自動翻訳」生来の訳語選定のランダム性が見られる旨を述べたが、同一の学習データを用いた特許 NT でも同様のランダム的な訳語変化が継続していることが確認されたことにより、この見解が妥当であることが裏付けられた。

「学员」「学習课件」「模拟练习」の三語においても、両エンジンで安定的に採用される最有力訳語は存在するが、その一方で例えば「学習课件」であれば第 15 回後の汎用 NT で「学習ウィジェット」、第 19 回後の特許 NT で「**学習記述**」といったマイナーな訳語が突発的に採用される回も依然として存在する。これらの用語においても、突発的な不適訳語の採用が完全に解消されたわけではないことがわかる。

例えば第 20 回後に特許 NT のみで初めて採用された、「模拟练习」の「练习」に相当する「エクササイズ」という訳語に関しても、詳しく調査したところ、第 15 回学習のタイミン

グで初めて13文の訳例がコーパス化されている。全コーパス中、「エクササイズ」の訳例はこの13文のみであることから、このタイミングで初めて「練習」の訳語候補のひとつに加わったことになる。本例ではその後も訳例数(64文)で優る「練習」が安定的に採用され続けており、引き続きこの訳語が最有力候補であることは確実だが、たとえ有力な訳語候補が存在する状況であっても、「エクササイズ」のようなマイナーな訳語候補も、学習データ中に訳例が存在する以上は、突発的に採用される可能性は皆無ではないということである。

とはいえ、全体的に見れば大規模アップデート以降、技術用語においてマイナーな訳語が突発的に採用される傾向はかなり弱まり、学習データ中で最有力の訳語候補が安定的に採用される確率が大幅に高まったことが本例の各技術用語の訳語の変遷からも見て取れる。技術用語の場合、中国審決コーパスデータでの網羅性は低く、コーパス中で最頻出な訳語が一般的にも最有力訳語であるとは限らない。例えば本例の「学员」や「模擬練習」に関しても、コーパス由来の「訓練生」や「シミュレーション練習」という訳語が、それ以前に度々採用されていた「学習者」や「シミュレーション演習」、正解訳文で採用されている「学生」や「模擬練習」よりも優れているとはいえない。しかしながら、少なくとも学習データ中で一定の出現頻度を有する訳語が安定的に採用されることにより、学習初期のように「学员：オペレータ」や「学员：教学者」、「学习课件：学習部」、「模擬練習：アナログトレーニング」といったマイナーな不適訳語がランダム的に採用されるリスクが少なくなり、その結果、全体的な翻訳品質が改善されたものと推察される。

なお、両エンジンの基礎学習データには異なる部分があるが、本例の各技術用語の訳語を見る限り、特に大規模アップデート以降は両エンジンで採用される訳語はほぼ共通していた。第3.3.3.3項では「両エンジンの各分野の技術用語の翻訳品質に、基礎学習データの違いによる顕著な差異は生じていない」と結論したが、本例の状況からも、技術用語のカバー状況に関して両エンジン間に顕著な差がないことがうかがえた。

以上、本例に関しても技術用語の訳語変化に焦点を絞って分析した。その結果、以下のことが確認できた。

- ・ 技術用語は、一般用語や文構造と同様、大規模アップデート以前はランダム的に変化する傾向が強く、学習データ中の出現頻度と直結しないマイナーな不適訳語が突発的に採用されることも少なくなかった。大規模アップデート以降はこの傾向が改善され、学習データ中で出現頻度の高い訳語が順当に採用される確率が高まったことが、本例における「学员：訓練生」や「模擬練習：シミュレーション練習」という技術用語の定点観測によって検証された。

この傾向の変化により、特殊な訳例に由来するマイナーな不適訳語が採用されるリス

クが軽減され、全体的な翻訳品質の底上げが果たされたと考えられる。

- また、このように学習データに準拠した訳語選定の傾向に変化したことで、大規模アップデート以降は、学習データに共通する部分の多い両エンジン間、すなわち汎用 NT と特許 NT の間で同一の訳語が採用される確率が高まり、各文においてエンジン間の翻訳品質の均質化が進行したと考えられる。
- ただし、本例の「匹配的」のように、学習データ中に同等に有力な訳語候補が複数存在する（と考えられる）語については、大規模アップデート以降もこれらの候補から毎回ランダム的に訳語が選定される傾向が残存していることも判明した。
- 本例で取り上げた各技術用語とも、特に大規模アップデート以降、頻繁に採用される有力な訳語候補は両エンジンで一致していた。両エンジンの基礎学習データには異なる部分があるが、本例の状況から、その差異は各エンジンで採用される技術用語の訳語に極端な違いをもたらすほどではないことがうかがえた。この結果は、第 3.3.3.3.項に記した「両エンジンの各分野の技術用語の翻訳品質に、基礎学習データの違いによる顕著な差異は生じていない」という結論とも整合している。

3.4.5. 実例 3-5 : 文番号#1155、汎用 NT

文番号#1155 は、無効宣告請求の書誌事項を一文にまとめて述べている文である。中国審決冒頭の「経緯」部分において、どの案件でもおおむね類似のフォーマットで必ず言及される。つまり中国審決における頻出文パターンに属する。まずは原文と正解訳文、無学習時の機械翻訳文を示す。

【文番号：#1155、汎用 NT】

原文	本无效宣告请求涉及专利号为 00131800.4、名称为“墨盒”的中国发明专利（下称本专利），本专利是 99800780.3 号发明专利申请的 分案申请 ，其申请日为 1999 年 05 月 18 日，最早优先权日为 1998 年 05 月 18 日，授权公告日为 2004 年 06 月 23 日，专利权人为精工爱普生株式会社。	
正解訳文	本無効宣告請求は出願番号第 0 0 1 3 1 8 0 0 . 4、名称「インクカートリッジ」の特許（以下、本特許という）に関するものであり、本特許は 9 9 8 0 0 7 8 0 . 3 の特許出願の 分割出願 であって、出願日は 1 9 9 9 年 5 月 1 8 日、最先の優先日は 1 9 9 8 年 5 月 1 8 日、登録日は 2 0 0 4 年 6 月 2 3 日、特許権者はセイコーエプソン株式会社である。	
無学習	本発明は 特許番号 が 00131800.4、名称が「インクカセット」である 中国発明特許 （以下本特許と称する）、本特許は 99800780.3 号 発明特許出願の請求申請 であり、その 申請日 は 1999 年 5 月 18 日、 最優先優先権日 は 1998 年 5 月 18 日、 許可公告日 は 2004 年 06 月 23 日、特許権者はセイコーエプソン株式会社である。	
	—	BLEU : 0.3782 (-) RIBES : 0.8702 (-)

本例には多数の特許・審決用語が含まれているが、無学習時の機械翻訳文では、赤字で示したとおり、その多くが誤訳・不適訳となっている。

特に問題なのは冒頭部で、原文の「**本无效宣告请求**」は正解訳文にあるとおり「**本無効宣告請求**」と訳すべき語であるが、無学習時の翻訳文はこれを「**本発明**」と、全く意味の異なる語に訳している。さらに中盤の「**分案申请**」も「**分割出願**」と訳すべきところ「**請求申請**」と、前半の「**分案**」に相当する意味が完全に失われ、代わりに原語とは無関係な「**請求**」という訳語が出力されている。

「**无效宣告请求**」「**分案申请**」は中国審決では多用される語だが、審決文書以外で使用されることはまれであり、基礎学習データ中には十分な訳例が存在していなかった可能性が高い。とはいえ、通常はそのような場合でも語を分解して訳出しているケースが多く、評価用コーパス中でこの語を含む他の文では、無学習時であっても、例えば前者は「無効

の宣告請求」(#826)、後者は「サブプランの申請」(#111) など、ナンセンスな訳語となる場合もあるにせよ、原語の直訳調の訳語が出力されている。

だが本例の訳語「**本発明**」「**請求申請**」は、いずれも原語とは全く意味が異なり、通常の誤訳の範疇を超えている。先に実例 3-3 で「解乳化ステップ」が「乳化**重合**ステップ」と誤訳されたことに関して、「みんなの自動翻訳」では何らかの理由で翻訳文から訳語が欠落した場合、その埋め合わせとしてその位置に文中の別の語句を重複出力するタイプの湧き出しエラーが発生する傾向があると述べたが、本例も「**発明**」に相当する「**发明**」、「**請求**」に相当する「**请求**」とも原文中に存在しており、同タイプの「欠落+湧き出し」エラーが発生したものと考えられる。

つまり本例では、他の文ではひとまず訳出できている「**无效宣告请求**」「**分案申请**」という語（の一部）が訳出できず欠落し、埋め合わせとして文中の別の語が湧き出している。このように、同じ用語であっても文によって訳出されたり、欠落して別の語で埋め合わされたりと扱いが変化するのには、ニューラル翻訳における翻訳文生成の方式が関係していると考えられる。

既に第 3.4.1.項で論じたとおり、ニューラル翻訳では、学習データ中の各用語について、各文例で用いられている訳語を、その原文における周辺情報と関連付けて多次元ベクトル化している。そして、その用語を含む文を機械翻訳する際は、各訳語に付されたベクトルと入力文の状況とを照合して、最もフィットすると判断した訳語を採用している。このため同じ用語であっても、入力文によって異なる訳語が採用される場合がある。そして本例の状況からは、入力文におけるその用語の状況（周辺にどのような語が存在するか等）に対し、採用に足る十分なベクトルを持つ訳語が学習データから得られない場合は、その用語の訳出自体がされないという可能性がある。さらに本例の状況からは、少なくとも当時（無学習時）の「みんなの自動翻訳」には、訳出できない用語が存在した場合、その欠落を文中の別の語句などで埋め合わせる性質があるものと見受けられる。

なお、本例では他にも「最早優先権日（最先の優先日）」が「**最優先優先権日**」と、やはり原語とニュアンスが異なる訳語が部分的に採用されている。「最早優先権日」の「最早」の訳語として採用された「**最優先**」は、全く無関係とも言い切れないが、単なる訳語選定のミスというには意味の乖離が大きく、直後の「**優先権日**」の「**優先**」が埋め合わせ的に重複出力されている可能性が高い。「最早」という語は単独では審決用語というほどの専門性はなく、コーパスを学習せずとも訳出できそうな一般語だが、ニューラル翻訳方式の場合、前述のとおり、学習データ中に「最早」の訳例があっても、この語が「優先権日」の周辺で用いられている文例が存在しないと、前例のない組合せの語のベクトルを

計算することになり、採用に足る十分なベクトルを持つ訳語が存在しない状況になりえる。おそらく本例はこれに該当し、かろうじて「最」は訳出できたが、「早」が訳出できず、その埋め合わせに直後の「優先」という語句を重複出力したのではないかと推察される。

こうしたニューラル翻訳方式における訳語決定のメカニズム、そして本例における「無効宣告請求」「分案申請」そして「最早優先権」の状況から、「みんなの自動翻訳」では、学習データ中に十分な訳例が存在しない用語は欠落しやすいと考えるべきである。そして無学習時においては、特許・審決用語の多くがこの「学習データ中に十分な訳例が存在しない用語」に該当する。このため、多くの文において本例に見られるような特許・審決用語関連の欠落が発生していた可能性は高い。

これに対し、中国審決コーパスデータ初回学習後の本例の翻訳文は以下のように変化した。

第1回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の中国特許(以下、本実用新案という)に関し、本実用新案は 99800780.3 号の特許出願の分割出願であり、出願日は 1999 年 5 月 18 日であり、最先の優先日は 1998 年 5 月 18 日であり、登録日は 2004 年 6 月 23 日であり、実用新案権者はセイコーエプソン株式会社である。	
	良化	BLEU : 0.6403 (+0.2621) RIBES : 0.9478 (+0.0776)

コーパスの学習により、無学習時に誤訳・不適訳であった特許・審決用語は全て適訳語に改善した。無学習時に訳語が欠落し、文中の別の語句で埋め合わされた結果と見なされた「本発明」と「請求申請」、「最優先優先権日」もそれぞれ「本無効宣告請求」「分割出願」「最先の優先日」と、いずれも正解訳文どおりの訳語が採用されるようになった。コーパス中の豊富な文例に基づき、これらの適訳語が新たに学習されたことで、採用に足る十分なベクトルを持つ訳語が得られた結果であろう。

ただしその一方で悪化した部分もあり、無学習時は正しく「本特許」「特許権者」と訳されていた箇所が「本実用新案」「実用新案権者」と誤訳されるようになり、原出願が実用新案と誤解される翻訳文となってしまっている。この誤訳は、原文の「專利」という語が、中国審決コーパスデータ中で「特許」と「実用新案」とに訳し分けられているために発生したものと考えられる。

「專利」は、中国における特許、実用新案、意匠の総称であり、対応する日本語が存在し

ない。このため本事業で作成した中国審決コーパスデータ中では、各案件の内容に応じて「特許」「実用新案」と訳し分けた。このコーパスを学習したことで、「專利」の訳語として「特許」に加えて「実用新案」という訳語候補が新たに学習されたと考えられる²⁵。この「專利」の訳し分けという課題に関しては、第 4.5.5.1.項及び第 5.5.項において再度考察する。

初回学習前後の翻訳品質を比較すると、この「專利」の訳し分け誤りが生じたにせよ、それ以外の特許・審決用語が全て適訳語に改善し、無学習時に発生していた重大な欠落も解消したため、相対評価では大幅な「**良化**」とみなせる。BLEU、RIBES のスコアも大幅に向上しており、学習前後の相対評価としては妥当である。

続いて、第 2 回学習後から第 5 回学習後までの翻訳文を示す。

【文番号：#1155、汎用 NT】

第 2 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の中国特許(以下、 本実用新案 という)に関し、 本実用新案 は 99800780.3 号特許出願の分割出願であり、その出願日は 1999 年 5 月 18 日、最先の優先日は 1998 年 5 月 18 日、登録日は 2004 年 6 月 23 日、 実用新案権者 はセイコーエプソン株式会社である。	
	同等	BLEU : 0.7164 (+0.0761) RIBES : 0.9604 (+0.0126)
第 3 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の中国特許(以下、 本特許 という)に関し、 本特許 は 99800780.3 号特許出願の分割出願であり、その出願日は 1999 年 5 月 18 日であり、最先の優先日は 1998 年 5 月 18 日であり、登録日は 2004 年 6 月 23 日であり、 特許権者 はセイコーエプソン株式会社であり、 セイコーエプソン株式会社 である。	
	良化	BLEU : 0.6775 (-0.0389) RIBES : 0.9566 (-0.0038)
第 4 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の中国特許(以下、 本特許 という)に関するものであり、 本特許 は 99800780.3 号の特許出願の分割出願であり、その出願日は 1999 年 5 月 18 日、最先の優先日は 1998 年 5 月 18 日、登録日は 2004 年 6 月 23 日、 特許権者 はセイコーエプソン株式会社である。	
	良化	BLEU : 0.8651 (+0.1876) RIBES : 0.9846 (+0.0280)

²⁵ デフォルト状態の「みんなの自動翻訳」では「專利」は「特許」とのみ訳されており、「実用新案」と訳されるケースは皆無であった。このため、デフォルト学習データ中には「專利」が「実用新案」と訳されている文例は存在しない可能性が高い。

第 5 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の中国特許(以下、 本特許 という)に関し、 本特許 は 99800780.3 号の特許出願の分割出願であり、その出願日は 1999 年 5 月 18 日であり、最先の優先日は 1998 年 5 月 18 日であり、登録日は 2004 年 6 月 23 日であり、 特許権者 はセイコーエプソン株式会社である。	
	同等	BLEU : 0.7216 (-0.1435) RIBES : 0.9631 (-0.0215)

上掲のとおり、初回学習後に適訳語に改善された各特許・審決用語はこの期間も常時安定して適訳語が採用されている。以降、最終学習回まで、わずかな例外²⁶を除いてこの状況は変わらなかった。つまり本例においても、他の実例と同様、特許・審決用語は学習初期から常時適訳語で安定した。

実例 3-1 の考察 (⇒3.4.1.) として、大規模アップデート以前の「みんなの自動翻訳」では各訳語に付された多次元ベクトルが学習データ追加の都度やや過剰に変化する傾向にあり、これが技術／一般用語のランダム的な訳語変化を生んでいたとの仮説を述べた。それにもかかわらず特許・審決用語では訳語の変化が見られなかったのは、こうした用語の多くは中国審決コーパスデータ上にしか訳例が存在せず、かつコーパスでは訳語が統一されていたためと考えられる。学習データ中に他の訳語候補が存在しないので、仮に訳語のベクトルが大きく変化しても、他に選択の余地がなく、訳語が変化しなかったと推察される。

これは、本例において「專利」の訳し分け誤りが発生した理由を説明するものでもある。「專利」の場合、コーパス上に「特許」と「実用新案」という二つの有力な訳語候補が存在している。この状況と、各回でベクトルが大きく変化する「みんなの自動翻訳」の特徴とにより、ランダム的な訳語変化が生じやすい状況になっていると考えられる。事実、「專利」は第 2 回学習後の翻訳文では引き続き「**本実用新案**」「**実用新案権者**」と訳されていたが、第 3 回後から第 5 回後にかけては再び「**本特許**」「**特許権者**」と正しい訳語に変化している。

この期間の翻訳文における「專利」以外の変化は、第 2 回後で「出願日は」に「その出願日は」と指示代名詞が付加されたこと、第 3 回後で後半の日付列挙が全て「…であり、…であり」と接続されたこと、第 4 回後で「…に関し、」が「…に関するものであり、」となったことなど、同義の範囲の言い回しの変化にとどまっており、実質的な内容に影響するものはほとんど存在しない。唯一、第 3 回後の文末で、なぜか「特許権者はセイコーエプソン株式会社であり、**セイコーエプソン株式会社である。**」と、同じ文言が繰り返された。代わりに欠落している情報はなく、内容理解上の問題はないが、日本語表現としては非常に不自然なエ

²⁶ 第 12 回時のみ「最早優先権日」が「最先の優先日」ではなく「最も早い優先日」と訳された。

ラーではある。ただし直後の第5回後で直ちに解消され、以降も再発していない。

この期間の各回の相対評価は、第2回学習後は実質的な変化がなく「同等」、これに対し第3回後は「専利」の訳し分け誤りが解消したため、文末の「セイコーエプソン株式会社」の繰り返しという不備が生じてはいるが差し引きで若干の「良化」、第4回後はこの不備も解消され、一切の不備がなくなったため更に「良化」、第5回後も実質的な変化はなく「同等」となる。これに対し、自動評価スコアは表層的な言い回しが一部変わっただけの第2回後に大きく向上していたり、実質的な変化がなかった第5回後には大きく低下していたりと、実態に即していない回が散見される。

続いて第6回学習後から第10回学習後までの翻訳文を示す。

【文番号：#1155、汎用 NT】

第6回学習後	本無効宣告請求は出願番号第00131800.4号、名称「インクカートリッジ」の中国特許(以下、 本特許 という)に関し、 本特許 は99800780.3号特許出願の分割出願であり、その出願日は1999年5月18日、最先の優先日は1998年5月18日、登録日は2004年6月23日、 特許権者 はセイコーエプソン株式会社である。
	同等 BLEU : 0.8179 (+0.0963) RIBES : 0.9754 (+0.0123)
第7回学習後	本無効宣告請求は出願番号00131800.4、名称「インクカートリッジ」の中国特許(以下、 本特許 という)に関し、 本特許 は99800780.3号特許出願の分割出願であり、その出願日は1999年5月18日であり、最先の優先日は1998年5月18日であり、 発行日 は2004年6月23日であり、 特許権者 はセイコーエプソン株式会社である。
	悪化 BLEU : 0.6960 (-0.1219) RIBES : 0.9560 (-0.0194)
第8回学習後	本無効宣告請求は出願番号00131800.4、名称「インクカートリッジ」の中国特許(以下、 本実用新案 という)に関し、 本実用新案 は99800780.3号の特許出願の分割出願であり、その出願日は1999年5月18日、最先の優先日は1998年5月18日、登録日は2004年6月23日、 実用新案権者 はセイコーエプソン株式会社である。
	悪化 BLEU : 0.7201 (+0.0241) RIBES : 0.9603 (+0.0043)
第9回学習後	本無効宣告請求は出願番号00131800.4、名称「インクカートリッジ」の中国特許(以下、 本実用新案 という)に関し、 本実用新案 は99800780.3号の特許出願の分割出願であり、その出願日は1999年5月18日、最先の優先日は1998年5月18日、登録日は2004年6月23日、 実用新案権者 はセイコーエプソン株式会社である。

	同等	BLEU : 0.7201 (±0) RIBES : 0.9603 (±0)
第 10 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の中国特許(以下、 本特許 という)に関し、 本特許 は 99800780.3 号特許出願の分割出願であり、その出願日は 1999 年 5 月 18 日であり、最先の優先日は 1998 年 5 月 18 日であり、登録日は 2004 年 6 月 23 日であり、 特許権者 はセイコーエプソン株式会社である。	
	良化	BLEU : 0.7181 (-0.0020) RIBES : 0.9627 (+0.0024)

第 6 回～第 10 回学習後の翻訳文の変遷も、おおむね第 5 回後までの延長線上にある。「専利」の訳し分け誤りは第 8 回後と第 9 回後にかけて再発しており、前記した推論どおりのランダム的な訳語変化が見られる。

それ以外の箇所の変化は第 5 回学習後までと同様のものが多く、その大多数は同義の範囲内での軽微な変化にとどまった。唯一、第 7 回後で特許用語「登録日」が「発行日」と誤訳され、翻訳品質の「悪化」の原因となったが、第 4 回後の「**セイコーエプソン株式会社**」の重複と同様、突発的な変化であり、この回のみで解消し、以降再発していない。

実質的な翻訳品質の相対評価は、第 6 回後は前回と「同等」の品質を維持したが、第 7 回後は「登録日」が「発行日」と誤訳されたため「悪化」、第 8 回後はこの突発的な誤訳は解消したが、再び「専利」の訳し分け誤りが発生したため差し引きではさらに「悪化」とした。第 9 回後は前回と完全一致の翻訳文につき「同等」、そして第 10 回後は訳し分け誤りが解消し、不備のない翻訳文に復帰したため「良化」と判定した。これに対し自動評価スコアは、「専利」の訳し分け誤りが再発した第 8 回後に両指標ともスコアが向上しているなど、引き続き実態と矛盾する回が見られる。

続いて第 11 回学習後から第 15 回学習後の翻訳文を示す。

【文番号：#1155、汎用 NT】

第 11 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の中国特許(以下、 本特許 という)に関し、 本特許 は 99800780.3 号特許出願の分割出願であり、その出願日は 1999 年 5 月 18 日であり、最先の優先日は 1998 年 5 月 18 日であり、登録日は 2004 年 6 月 23 日であり、 特許権者 はセイコーエプソン株式会社である。	
	同等	BLEU : 0.7181 (±0) RIBES : 0.9627 (±0)
第 12 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の中国特許(以下、 本特許 という)に関し、 本特許 は第 99800780.3 号の特	

	許出願の分割出願であり,その出願日は 1999 年 5 月 18 日,最も早い優先日は 1998 年 5 月 18 日,登録日は 2004 年 6 月 23 日,特許権者はセイコーエプソン株式会社である。
	悪化 BLEU : 0.7523 (+0.0342) RIBES : 0.9623 (-0.0004)
第 13 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の特許(以下、本特許という)に関し,本特許は第 99800780.3 号の特許出願の分割出願であり,その出願日は 1999 年 5 月 18 日,最先の優先日は 1998 年 5 月 18 日,登録日は 2004 年 6 月 23 日,特許権者はセイコーエプソン株式会社である。
	良化 BLEU : 0.8310 (+0.0787) RIBES : 0.9732 (+0.0109)
第 14 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の特許(以下、本特許という)に関し,本特許は第 99800780.3 号の特許出願の分割出願であり,その出願日は 1999 年 5 月 18 日,最先の優先日は 1998 年 5 月 18 日,登録日は 2004 年 6 月 23 日,特許権者はセイコーエプソン株式会社である。
	同等 BLEU : 0.8310 (±0) RIBES : 0.9732 (±0)
第 15 回学習後	本無効宣告請求は出願番号 00131800.4,名称「インクカートリッジ」の特許(以下、本特許という)に関し,本特許は第 99800780.3 号特許出願の分割出願であり,その出願日は 1999 年 5 月 18 日,最先の優先日は 1998 年 5 月 18 日,登録日は 2004 年 6 月 23 日,特許権者はセイコーエプソン株式会社である。
	同等 BLEU : 0.7961 (-0.0349) RIBES : 0.9688 (-0.0044)

第 12 回学習後以降は「みんなの自動翻訳」の大規模アップデート後の翻訳文となる。大規模アップデート以降、「特許」は最終学習後まで毎回正しく「特許」と訳されるようになった。この状況は、大規模アップデートを機にベクトル付与のロジックが改良され、過度の変動が抑制されたとする本章での考察内容と整合する。ただし、他の文では大規模アップデート以降も「特許」の訳語変化は少なからず検出されており（第 4.5.5.1.項において後述）、訳し分け問題が根本的に解決したわけではない。

第 11 回後～第 15 回後における特記事項としては、これまで安定して「最先の優先日」と訳されていた審決用語「最早優先権日」が、第 12 回後のみ「最も早い優先日」と異なる訳語で訳されたことが挙げられる。同義であり内容理解上は問題ないが、第 7 回後の「発行日」同様、特許・審決用語であっても突発的な訳語変化が生じる可能性は皆無ではないことが示された。

この期間は、第 12 回後の「最も早い優先日」への変化による軽微な「悪化」を除けば、各回とも言い回しが微妙に異なるのみで品質は完全に同等であった。これに対し、自動評価スコアは特に BLEU で大きな変動が生じているのが目立つ。

最後に、第 16 回学習後から最終第 20 回学習後の翻訳文を示す。

【文番号：#1155、汎用 NT】

第 16 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の特許(以下、 本特許 という)に関し、 本特許 は 99800780.3 号特許出願の分割出願であり、その出願日は 1999 年 5 月 18 日、最先の優先日は 1998 年 5 月 18 日、登録日は 2004 年 6 月 23 日であり、 特許権者 はセイコーエプソン株式会社である。
	同等 BLEU : 0.8080 (+0.0119) RIBES : 0.9749 (+0.0061)
第 17 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の特許(以下、 本特許 という)に関し、 本特許 は第 99800780.3 号の特許出願の分割出願であり、その出願日は 1999 年 5 月 18 日、最先の優先日は 1998 年 5 月 18 日、登録日は 2004 年 6 月 23 日、 特許権者 はセイコーエプソン株式会社である。
	同等 BLEU : 0.8310 (+0.0230) RIBES : 0.9732 (-0.0017)
第 18 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の特許(以下、 本特許 という)に関し、 本特許 は第 99800780.3 号の特許出願の分割出願であり、その出願日は 1999 年 5 月 18 日、最先の優先日は 1998 年 5 月 18 日、登録日は 2004 年 6 月 23 日、 特許権者 はセイコーエプソン株式会社である。
	同等 BLEU : 0.8310 (±0) RIBES : 0.9732 (±0)
第 19 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の特許(以下、 本特許 という)に関し、 本特許 は第 99800780.3 号の特許出願の分割出願であり、その出願日は 1999 年 5 月 18 日、最先の優先日は 1998 年 5 月 18 日、登録日は 2004 年 6 月 23 日、 特許権者 はセイコーエプソン株式会社である。
	同等 BLEU : 0.8310 (±0) RIBES : 0.9732 (±0)
第 20 回学習後	本無効宣告請求は出願番号 00131800.4、名称「インクカートリッジ」の特許(以下、 本特許 という)に関し、 本特許 は第 99800780.3 号特許出願の分割出願であり、その出願日は 1999 年 5 月 18 日、最先の優先日は 1998 年 5 月 18 日、登録日は 2004 年 6 月 23 日、 特許権者 はセイコーエプソン株式会社である。

	同等	BLEU : 0.8184 (-0.0126) RIBES : 0.9721 (-0.0011)
--	----	--

第 16 回後から最終第 20 回後にかけても、大規模アップデート後の機械翻訳文の傾向がそのまま継続している。「專利」は引き続き毎回「特許」と正しい訳語で安定し、第 12 回後に「最も早い優先日」と軽微な不適訳が突発的に採用されていた箇所も、その後は各回とも「最先の優先日」という適訳語が採用されている。

また、第 17 回後から第 19 回後にかけては完全に同一の機械翻訳文が連続して出力されている。この翻訳文は第 13~14 回後とも同一である。翻訳文が変化した第 16 回後と第 20 回後も、変化が生じたのは「本特許は第 99800780.3 号の特許出願の分割出願であり」という一節のみであり、第 16 回後はこの一節から出願番号直前の「第」と直後の「の」が省略されて「本特許は 99800780.3 号特許出願の分割出願であり」となっているのみ、第 20 回後は上記「の」が省略されているのみで、いずれも内容は完全に同じである。つまり本例に関しても、ここまでの各実例で見られた学習前後における機械翻訳文の安定化、そして翻訳文が変化した場合もその変化は小規模化、局所化し、かつ同義の範囲内に収まるという大規模アップデート以降の「みんなの自動翻訳」の傾向が顕著に表れている。

本例の定点観測からは、以下のことが確認できた。

- ・ 本例においても、特許・審決用語の多くは無学習時には正しく訳出できていなかったが、中国審決コーパスデータの学習により各語とも早期に適訳語に改善され、以降も最終学習時までほぼ常時安定した。
- ・ ただし唯一、日本語の「特許」「実用新案」「意匠」の総称である「專利」という語のみは、案件によって適切に訳し分ける必要があり、本例では「特許」と訳するのが適切であったが、これを「実用新案」と誤訳している回が散見された。
- ・ 本例における「專利」の訳し分け誤りは、中国審決コーパスデータ中に有力な訳語候補が「特許」と「実用新案」の二種類存在することと、ニューラル機械翻訳における訳語選定の方法論とに起因する。本例では大規模アップデートを境に「專利」の訳し分け誤りは発生しなくなったが、他の文では引き続き発生しているものもあり、この課題は根本的には解決していない。

3.4.6. 実例 3-6 : 文番号#547、汎用 NT [翻訳不能文とその改善]

本調査分析に使用した評価用コーパスは、原則として案件単位に、その全文を採用している (⇒2.4.1.)。このためコーパスに含まれる文長は広範囲に及んでいるが、調査開始当初、評価用コーパスとして選定したうちの数文について、文長が長すぎて「みんなの自動翻訳」では翻訳されないことが判明した (⇒3.3.1. 【付記】 参照)。

具体的には評価用コーパス中の 19 文²⁷がこれに該当しており、これらの文は機械翻訳文が取得できず、BLEU、RIBES のスコアも算出されなかったことから、自動評価スコアの集計対象から除外せざるを得なかった。このため本報告書に示した自動評価の全件平均スコアは全て、評価用コーパス 2,013 文からこれら 19 文を除外した 1,994 文の平均値を用いている。

その後、これら翻訳不能文 19 文は、第 5 回学習後の機械翻訳処理 (2020 年 1 月) で初めて機械翻訳文が出力され、以降は最終学習後まで、ごく一部の例外²⁸を除けば、各回とも機械翻訳文が取得できた。本項では、これら翻訳不能文から文番号#547 を実例に採り、その改善状況と、残存する課題について論じる。

まずは文番号#547 の原文と正解訳文を以下に示す。

【文番号 : #547、汎用 NT】

原文	经合议审查, 合议组于 2017 年 05 月 26 日向复审请求人发出复审通知书, 指出: 由于权利要求 1-3 要求保护的是鼻腔插入管用固定件, 并不包括鼻腔插入管, 因此对比文件 1 实际上已经公开了权利要求 1 的技术方案, 权利要求 1 不具备专利法第二十二条第二款规定的新颖性; 权利要求 2 和 3 的附加特征或者被对比文件 9 公开, 或者是本领域的常规技术手段, 因而不具备专利法第二十二条第三款规定的创造性; 对于权利要求 4 和 5 请求保护的包括固定件和安装在固定件的可插入至咽部的鼻腔插入管所构成的鼻腔插入设备组, 由于对比文件 1 中插入鼻腔的过滤器与对比文件 7 中插入至鼻腔深处的管二者都是可插入至鼻腔的管体, 一般情况下其固定件可相互适应是合理的, 对比文件 7 也给出了在管较长的情况下将固定件固定在管的基端部的技术启示, 至于对比文件 1 中的 U 型桥构件, 其并不是只起到连接或桥接作用, 也发挥将连接或桥接对象稳
----	--

²⁷ 具体的には、文番号#223、#340、#406、#413、#547、#578、#695、#964、#994、#1067、#1122、#1143、#1145、#1200、#1266、#1344、#1666、#1888、#1898 の 19 文である。

²⁸ 具体的には、汎用 NT の第 10 回学習後の文番号#1122 と、同じく汎用 NT の第 11 回学習後の文番号 944 のみ、機械翻訳結果を取得できなかった。

	定固定的作用, 因此, 本领域技术人员具有将上述对比文件进行结合的动机, 从而权利要求 4 和 5 不具备专利法第二十二条第三款规定的创造性。
正解訳文	<p>審理を経て, 合議体は 2017 年 5 月 26 日に復審請求人へ復審通知書を送付し, 以下のように指摘した: 請求項 1~3 が特許請求するものは鼻腔挿入チューブ用固定具であり, 鼻腔挿入チューブは含まれておらず, よって引用文献 1 には, 請求項 1 の技術的解決手段が実際に開示されているため, 請求項 1 は特許法第 22 条第 2 款に規定される新規性を有さない; 請求項 2 及び 3 の付加的な特徴は, 引用文献 9 によって開示されているか, または本分野における一般的な技術手段であるため, 特許法第 22 条第 3 款に規定される創造性を有さない; 請求項 4 及び 5 が特許請求しているものは固定具及び固定具に取り付けられる咽頭部に挿入可能な鼻腔挿入チューブで構成される鼻腔挿入デバイスセットであって, 引用文献 1 中の鼻孔に挿入するフィルタと, 引用文献 7 中の鼻腔内の奥に挿入するチューブの両者はいずれも鼻腔に挿入するチューブ体であり, 一般的にその固定具は相互に適用可能であると見なすのが適当であり, 引用文献 7 においてもチューブが長い場合に, 固定具をチューブの基端部に固定するという技術的示唆が示されており, 引用文献 1 中の U 型ブリッジ部材については, それは連結または架橋の役割を果たすだけでなく, 連結または架橋の対象を安定して固定する役割も果たし, このため, 当業者には上記引用文献を組み合わせる動機があり, よって請求項 4 及び 5 は特許法第 22 条第 3 款に規定される創造性を有さない。</p>

本例は中国語原文で 442 文字に及ぶ長文である。ただしその実態は 4 つの文がコロンやセミコロンで接続されたものであり、正解訳文に下線で示したとおり、第 1 文が「…指摘した:」(2 行目) まで、第 2 文が「…新規性を有さない;」(6 行目) まで、第 3 文が「…創造性を有さない;」(9~10 行目) まで、第 4 文が「…創造性を有さない。」(最終行) までとなっている。

当初、「みんなの自動翻訳」ではこの文は文長制限により翻訳不能であったが、前記したとおり、第 5 回学習後から機械翻訳結果が得られるようになった。以下、第 5 回学習後の機械訳文を示す。

第 5 回学習後	,請求項 1-3 が保護を請求しているのは鼻腔挿入管の固定部材であって鼻腔挿入管を含まないために,引用文献 1 は実際に請求項 1 の技術的解決手段を開示しており,請求項 1 は特許法第 22 条第 2 款に規定される
----------	---

	<p>新規性を有さない;請求項 2 と 3 の付加的な特徴は、引用文献 9 に開示されているか、又は本分野の一般的な技術手段であり、したがって、特許法第 22 条第 3 款に規定される<u>創造性を有さない</u>;請求項 4 と 5 で保護を請求している、固定部材と固定部材をチューブの基端部に挿入するという技術的示唆を与えており、引用文献 1 における、鼻腔に挿入されたフィルタとは、引用文献 7 において鼻腔の深さの位置に挿入されると、固定具をチューブの基端部に固定するという技術的示唆も与えており、引用文献における、鼻腔に挿入されるフィルタは、引用文献において、鼻腔の深さの位置に挿入された 2017 1 2 7 26 4 5 22 3。</p>
	<p>(BLEU : 0.2915 RIBES : 0.7272)</p>

機械翻訳文が出力されるようにはなかったが、上掲のとおり、第 5 回後の機械翻訳文は正解訳文に比べて明らかに文長が短く、具体的には正解訳文が 608 文字であるのに対して 391 文字しか出力されていない。文の構成を詳しく見ていくと、正解訳文の冒頭から「…指摘した:」までにあたる第 1 文が完全に欠落しており、いきなり「請求項 1 - 3 が…」という第 2 文から開始されている。この第 2 文と続く第 3 文はおおむね原文と対応した内容になっているが、第 4 文に関しては「挿入された」と、明らかに文の途中で翻訳が途絶しており、かつ係り受けが大幅に乱れていて意味を成していない。さらには、原文中の数値が文末にまとめて羅列されるというエラー²⁹も発生している。

このような大規模な欠落により、本例の第 5 回学習後の機械翻訳品質はきわめて低い。こうした状況は他の長文でも同様であり、例えば機械翻訳文の文末に数値が羅列された文は本例を含め 19 文中 17 文に及ぶ。脚注に記したとおり、文末の数値羅列は原文内容の大規模な欠落を示すものであり、つまりは本件と同様の状況が生じていると見なせる。

こうした状況は第 6 回学習後以降もしばらく続いたが、「みんなの自動翻訳」の大規模アップデートを経た第 12 回学習後以降は、機械翻訳文の傾向が明らかに変化した。以下、本例の第 12 回学習後の翻訳文を示す。

²⁹ 本例の機械翻訳文末尾には「2017 1 2 7 26 4 5 22 3」という数字の羅列がある。これは「みんなの自動翻訳」の典型的な数値エラーであり、詳しくは第 4 章「人手評価結果」の第 4.5.5.8.項で解説するが、簡略に述べれば「機械翻訳文で欠落した箇所に出力される予定であった数値が行き場を失い文末にまとめて出力される」エラーである。本例の場合は、第 1 文の「2017 年 5 月 26 日」と第 4 文後半の「引用文献 1」「引用文献 7」「請求項 4 及び 5」「特許法第 22 条第 3 款」に相当する数値が（順不同に）列挙されているようである。ただし若干の齟齬があり（5 が足りず 2 が余る）、完全には一致していない。

【文番号：#547、汎用 NT】

第 12 回学習後	<p>審理を経て、合議体は 2017 年 5 月 26 日に復審請求人へ復審通知書を送付し、以下を指摘した:請求項 1-3 が特許請求しているのは鼻腔挿入管用の固定具であり、鼻腔挿入管を含まないため、引用文献 1 は実質的に請求項 1 の技術的解決手段を開示しており、請求項 1 は特許法第 22 条第 2 款に規定される新規性を有さない;請求項 2 及び 3 の付加的特徴は引用文献 9 に開示されているか、又は本分野の一般的な技術手段であるため、特許法第 22 条第 3 款に規定される創造性を有さない;請求項 4 及び 5 が特許請求している固定具及び固定具に取り付けられた咽頭部に挿入可能な鼻腔挿入管で構成された鼻腔挿入デバイスセットについて、引用文献 1 における鼻腔に挿入されたフィルターと引用文献 7 における鼻腔の深さまで挿入された管とはいずれも鼻腔に挿入可能な管体であり、一般的にはその固定具が互いに適応できることは合理的であり、引用文献 7 も管が長い場合に固定具を管の基端部に固定するという技術的示唆を与えており、引用文献 1 における U 字型ブリッジ部材について、それは接続又はブリッジ接続の役割を果たすだけでなく、接続又はブリッジ接続対象を安定的に固定する役割も果たし、従って、当業者は上記引用文献を組み合わせる動機があり、従って請求項 4 及び 5 は特許法第 22 条第 3 款に規定される創造性を有さない。</p> <p>BLEU : 0.6132(+0.3217) RIBES : 0.9216 (+0.1944)</p>
-----------	---

第 5 回後とは異なり、第 12 回後の機械翻訳文では第 1 文から第 4 文まで全ての文が出力されている。文字数も 568 文字と正解訳文 (608 文字) に近づいた。文末の数字の羅列も解消しており、正解訳文と読み比べてみても、若干の日本語のたどたどしさや、採用された技術用語の不一致はあるものの、各文とも文意はほぼ完璧に把握でき、大幅な品質改善を遂げている。

他の長文も本例と同様、大規模アップデートを契機に明らかに出力文字数が増加している。19 文の大規模アップデート前後の出力文字数を比較すると、アップデート直前の第 11 回学習後の平均文字数が 373 文字であったのに対し、アップデートが実施された第 12 回後は 591 文字に増加しており (汎用 NT)、それ以降も同程度の文字数が出力されており、翻訳文の途絶や欠落への対処が講じられたことがわかる。

ただし、出力文字数が増加したとはいえ、依然として翻訳文が途絶しているものも少なくない。例えば最終学習後の汎用 NT では、19 文中 16 文で依然として翻訳文が途絶し、文が

完結していない。本例に関しても、第 12 回後は文末まで出力されていたが、第 20 回後では再び翻訳文が途絶した。以下に第 20 回後の翻訳文を示す。

【文番号：#547、汎用 NT】

第 20 回学習後	<p>合議体は審理を経て、2017 年 5 月 26 日に復審請求人へ復審通知書を送付し、以下を指摘した：請求項 1-3 が特許請求しているのは鼻腔挿入管用固定部材であり、鼻腔挿入管を含まないため、引用文献 1 は実際に請求項 1 の技術的解決手段を開示しており、請求項 1 は特許法第 22 条第 2 款に規定される新規性を有さない；請求項 2 及び 3 の付加的特徴は引用文献 9 に開示されているか、又は本分野の一般的な技術手段であるため、特許法第 22 条第 3 款に規定される創造性を有さない；請求項 4 及び 5 が特許請求している固定部材及び固定部材に取り付けられた咽頭部に挿入可能な鼻腔挿入管からなる鼻腔挿入装置セットについて、引用文献 1 における鼻腔に挿入するフィルターと引用文献 7 における鼻腔の奥まで挿入する管とはいずれも鼻腔に挿入可能な管体であり、一般的にその固定部材が互いに適応できることは合理的であり、引用文献 7 にも管が長い場合に固定部材を管の基端部に固定するという技術的示唆がなされており、引用文献 1 における U 字型ブリッジ部材について、それは接続又はブリッジ接続の作用のみを果たすものではなく、接続又はブリッジ接続対象を安定して固定する作用を果たすものであり、したがって、当業者には上記引用文献を組み合わせる動機があり、したがって請求項 4 及び 5 は特許</p>
	<p style="text-align: center;">BLEU : 0.5441(-0.0691) RIBES : 0.9029 (-0.0187) (第 12 回学習後との比較)</p>

上掲のとおり、本例の第 20 回学習後の翻訳文は文末で明らかに途絶している。第 12 回後は同じ文が文末まで、全 568 文字が出力されていたが、今回は 555 文字しか出力されていない。最終学習後に汎用 NT で翻訳文が途絶していた 16 文のうち、最も短いものは 447 文字で途絶していたが (#1145)、最も長いものでは翻訳文は 867 文字まで出力されていた (#1666)。また翻訳文が末尾まで出力された 6 文のうち最長のものは 511 文字 (#1344) であるなど、翻訳文が特定の文字数を越えた時点で途絶するという単純な事象ではないようである。原文を基準としても、例えば本項で取り上げた #547 は同じ原文であるにもかかわらず途絶した回としない回があり、第 20 回後の #1888 では汎用 NT では 519 文字で翻訳が途絶し、特許 NT では翻訳文が 526 文字で文末まで出力されているなど、途絶が発生する条件は判然としない。

このように発生条件は特定不可能であるが、原文が極端な長文である場合、「みんなの自動翻訳」では依然として翻訳文が途絶する可能性が高いことは確かであり、現状の課題のひとつとして挙げられる。

4. 人手評価結果とその分析

4.1. 人手評価の概要

人手評価は、自動評価と同様、「みんなの自動翻訳」の汎用 NT と特許 NT それぞれの機械翻訳結果を対象に実施した。

人手評価では、評価用コーパス 2,013 文のうち 100 文のみを対象とした。また、自動評価が中国審決コーパスデータを約 10 万文学習させる都度実施したのに対し、人手評価は約 50 万文の学習ごと、つまり無学習時、第 5 回学習後、第 10 回学習後、第 15 回学習後、第 20 回学習後の計 5 回実施した。

人手評価は 1 文あたりの作業負担が大きく大量の文を評価するには適さないが、正解訳文との一致度のみを基準とする自動評価に比べ、機械翻訳文の品質をより正確かつ精密に評価することができる。本調査分析ではこの利点を活かし、自動評価結果から看取された「みんなの自動翻訳」の機械翻訳品質の改善状況について、サンプリングによる人手評価の結果と照合することで、その妥当性を検証した。

さらに、評価対象とした個々の機械翻訳文においてどのような種類のエラーが発生しているかや、コーパスの学習によるこれらのエラーの解消の状況などについても目視で精密に調査し、中国審決という文書、並びに「みんなの自動翻訳」というシステムにおける機械翻訳品質の傾向や課題を詳細に分析した。

4.2. 人手評価観点

人手評価では、評価対象各文について下記 4 つの観点から評価を実施した。

- ① 内容伝達レベルの評価
- ② 個々の翻訳エンジンの学習前後の相対評価
- ③ 両エンジン間の機械翻訳文の相対評価
- ④ エラーのカテゴリ別カウント

4.3. 人手評価者

人手評価者には本事業の翻訳内容校閲者を起用した。文ごとに担当の評価者をあらかじめ定め、原則全ての回の評価を同一の者が担当した。

4.4. 人手評価の対象文

人手評価対象文には、評価用コーパスとして選定した 2,013 文から 100 文を選定した。

評価用コーパスは文献単位に全 31 文献から選定されている (⇒2.4.1.)。人手評価対象文は、これら 31 文献から均等に 3~4 文ずつ選定した。その際は、極端な短文や公報由来の文 (主に「特許請求の範囲」の摘記) は除外し、かつ中国審決に頻出の用語・表現をできる限り網羅した。

なお、選定した 100 文の平均文字数は 73.8 文字であり、中国審決コーパスデータ全体の平均文字数 (70.3 文字) とほぼ同等である。

4.5. 各評価項目の集計結果

本項において、全5回にわたり実施した人手評価の結果をまとめる。第4.2.項に示した各評価観点、すなわち①内容伝達レベルの評価、②個々の翻訳エンジンの学習前後の相対評価、③両エンジン間の機械翻訳文の相対評価、④エラーのカテゴリ別カウントのそれぞれについて、各回の集計結果を示すとともに、評価結果の分析・考察を行う。

4.5.1. 内容伝達レベルの評価

機械翻訳文の品質の絶対評価として、特許庁が提案する「特許文献機械翻訳の品質評価手順 (ver.1.0)」の主要な評価基準である「内容伝達レベル」の評価を行った。

内容伝達レベルの評価では、原文の実質的な内容が機械翻訳文からどの程度把握できるかを下記5段階の評価基準で主観的に判定した。

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

下表に、評価対象文100文の、翻訳エンジン別／学習段階別の内容伝達レベル（5～1の5段階評価）の集計結果を示す。

表 4.5.1. 「内容伝達レベル」 人手評価スコアの集計結果

エンジン	学習	平均	5	4	3	2	1
汎用 NT	無学習時	2.41	0	16	32	29	23
	第5回学習後	4.42	56	30	14	0	0
	第10回学習後	4.32	55	26	15	4	0
	第15回学習後	4.61	65	31	4	0	0
	第20回学習後	4.65	68	29	3	0	0
特許 NT	無学習時	2.18	2	9	26	31	32
	第5回学習後	4.28	49	35	12	3	1
	第10回学習後	4.30	56	27	8	9	0
	第15回学習後	4.66	68	30	2	0	0
	第20回学習後	4.61	64	33	3	0	0

中国審決コーパスデータの学習により、人手評価対象 100 文における内容伝達レベルの平均値は、汎用 NT では無学習時の 2.41 から最終学習後は 4.65 (+2.24) へ、特許 NT では無学習時の 2.18 から最終学習後は 4.61 (+2.43) へと、それぞれ大きく向上した。両システムとも無学習時から第 5 回学習後にかけて最も大幅に向上しており、汎用 NT では +2.01、特許 NT では +2.10 と約 2 レベル向上した。続く第 10 回学習後は汎用 NT が -0.10、特許 NT は +0.02 と微かな変動で事実上の横ばいとなったが、第 12 回学習前の「みんなの自動翻訳」大規模アップデート (⇒2.2.3.) を経た第 15 回学習後にはそれぞれ +0.29、+0.36 と再び明瞭な向上が見られた。その後、最終学習となる第 20 回学習後においても両エンジンの変動はそれぞれ前回比 +0.04、-0.05 と誤差レベルにとどまり、大規模アップデート直後に到達した水準が維持された。

人手評価は、評価用コーパス全件を対象に実施した自動評価の妥当性を検証し、より精緻な品質改善状況を把握するために実施している。以下、上述した内容伝達レベルの推移を自動評価スコアと照合し、自動評価結果から看取した品質改善状況の妥当性を検証するとともに、自動評価スコアでは漠然としていた、各学習段階で達成された具体的な機械翻訳品質の程度について把握する。

無学習時の内容伝達レベルは、汎用 NT が 2.41、特許 NT が 2.18 とともにレベル 2 周辺であった。内容伝達レベルのレベル 2 は「いくつかの重要情報は正確に伝達されている (20%～)」程度の品質、レベル 3 は「半分以上の重要情報は正確に伝達されている (50%～)」であることを示しており、デフォルト状態の「みんなの自動翻訳」での中国審決文の翻訳品質が、平均して「50%未満の重要情報しか伝達されない」レベルであることが具体的に示された。

これに対し、中国審決コーパスデータによる初期学習を経た第 5 回学習後の内容伝達レベルは、汎用 NT が 4.42、特許 NT が 4.28 とそれぞれレベル 4 周辺まで顕著に向上した。この結果は、初回学習後に顕著に向上し、第 5 回後もその水準を維持した自動評価スコアの学習初期段階の推移と整合しており、自動評価結果の妥当性を裏付けている。さらに、内容伝達レベル 4 の定義「ほとんどの重要情報は正確に伝達されている (80%～)」に照らし、第 5 回学習後の翻訳品質が平均して「80%以上の重要情報が伝達される」レベルにまで改善されたことが把握できる。

上表には評価対象文 100 文の内容伝達レベル 1～5 の分布も示した。無学習時の分布では、「すべての重要情報が正確に伝達されている (100%)」ことを示すレベル 5 は汎用 NT では皆無、特許 NT でも 2 文とごく僅かであり、品質がきわめて低いレベル 1～2 を付された文がそれぞれ 52 文、63 文と過半数を占めていた。これに対し第 5 回学習後の分布では、

レベル 5 が汎用 NT で 56 文、特許 NT で 49 文と半数前後を占め、レベル 2 以下の文は汎用 NT では皆無、特許 NT もレベル 2 が 4 文のみとごく少数で、ほぼ全文において一定以上の翻訳精度が達成されたことが示された。

続く第 10 回学習後は、汎用 NT は 4.32 (-0.10)、特許 NT が 4.30 (+0.02) と、両エンジンとも内容伝達レベルの平均値は前回時とほぼ同水準にとどまった。この期間、自動評価の全体平均スコアも学習初期に到達した水準を維持しつつ毎回微増と微減を繰り返す状況であり、両評価結果は引き続き整合した。第 3.4.項で定点観測した各実例の状況からは、第 5 回学習後の時点で既に大半の特許・審決用語は適訳語で訳されるようになっており、その適訳語が第 10 回学習後の時点でも各文で安定して採用された結果の「現状維持」と解釈される。これは裏を返せば、第 5 回学習後の時点で既に主要な特許・審決用語には改善の余地がほぼ無くなっており、それ以降の全体品質の変動は主として技術／一般用語の不安定な訳語変化によるものになっていると考えられる。

第 15 回学習後は、第 12 回学習前に実施された「みんなの自動翻訳」大規模アップデートを経た後の機械翻訳品質の評価となる。自動評価ではアップデート直後にスコア水準が明らかに向上したが、人手による内容伝達レベルの評価でも同様の結果が得られた。具体的には、両エンジンとも第 10 回後は第 5 回後から微動のレベル 4.3 近辺で停滞していたところ、アップデートを経た今回（第 15 回後）は汎用 NT が 4.61、特許 NT が 4.66 と、いずれもレベル 4.6 周辺まで上昇した。この結果から、自動評価結果が示していた大規模アップデートの翻訳品質改善効果が実体を伴うものであることが確認された。

最終回となる第 20 回) 学習後の人手評価結果は、本事業で作成し中国審決対訳データ全件を学習させることで達成される、いわば実用時の翻訳品質と仮定できる。両システムの全体平均レベルは汎用 NT が前回比+0.04 の 4.65、特許 NT は前回比-0.05 の 4.61 と、それぞれ微動にとどまり、前回時の品質水準がそのまま維持されていると見なせる。この結果により、前回時に向上した前回時の品質水準が偶然の偏りによる一時的なものではなく、大規模アップデート以降の「みんなの自動翻訳」で常時期待できる品質であることが確認された。

以上、人手評価による内容伝達レベルの集計結果について分析した。各回の結果は、第 3.3.1.項において自動評価結果が示した品質動向、すなわち学習初期において翻訳品質が大幅に向上し、その後しばらくは同水準上での微動にとどまったが、大規模アップデートを機にさらなる品質改善が果たされ、以降もその高水準が維持されたという推移と完全に整合するものであった。また、汎用 NT と特許 NT の両エンジンの内容伝達レベルもコーパス学習後は常時ほぼ同等であり、こちらも自動評価結果と一致した。

このように、評価用コーパス全件を対象とした自動評価結果と、人手による内容伝達レベルの評価結果とが各学習段階において完全に整合したことにより、本調査分析で実施した自動評価結果が十分な精度を有しており、これに基づき論じた第 3 章の分析・考察結果が十分な妥当性を有することが検証された。

また、人手による内容伝達レベルの評価を通じ、自動評価では明らかでなかった具体的な翻訳品質水準についても明瞭に把握することができた。無学習時の翻訳品質は両エンジンとも内容伝達レベルは平均でレベル 2 強と低く、個々の文単位でもレベル 5 に達していた文はほぼ皆無という状況であったのに対し、中国審決コーパスデータを学習させた第 20 回学習後の評価では、全体の 65%前後の文がレベル 5、残りの文も大部分がレベル 4 であり、両エンジンでレベル 4~5 の比率が 97%を占めるという、極めて優秀な翻訳品質に達していることが示された。

4.5.2. 個々の翻訳エンジンの学習前後の相対評価

内容伝達レベルの評価は5段階の基準に基づく絶対評価であるため、学習前後の機械翻訳文に有意な品質差があっても、常にレベルに差がつくとは限らない。このため本評価観点では、汎用NT、特許NTそれぞれのエンジン別に学習前後の機械翻訳結果の相対的な優劣を評価し、学習による改善効果の有無を評価した。

評価は、学習後の機械翻訳文の品質が学習前に比べて「向上」したか「劣化」したか、若しくは「同等」であるかの三者択一とし、評価者には、学習前後の翻訳文に何らかの差異が生じている場合、できるだけ優劣をつけるよう求めた。下表に各回における集計結果を示す。

表 4.5.2. 各回学習前後の翻訳品質の「相対評価」結果

翻訳エンジン	比較対象		学習後に		
	学習前	学習後	向上	同等	劣化
汎用 NT	無学習時	第 5 回学習後	95	4	1
	第 5 回学習後	第 10 回学習後	24	52	24
	第 10 回学習後	第 15 回学習後	44	43	13
	第 15 回学習後	第 20 回学習後	16	71	13
特許 NT	無学習時	第 5 回学習後	95	2	3
	第 5 回学習後	第 10 回学習後	29	51	20
	第 10 回学習後	第 15 回学習後	45	48	7
	第 15 回学習後	第 20 回学習後	9	81	10

無学習時と第 5 回学習後との比較では、両エンジンともほぼ全文が学習後に「向上」したと判定された。中国審決コーパスデータがここで初めて学習され、その結果、頻出の特許・審決用語の大半が適訳語に改善されていたことを考えれば、妥当な結果といえる。またこの結果は、無学習時の機械翻訳文のほぼ全件に何らかの改善の余地があったことを示している。

続く第 5 回学習後と第 10 回学習後との比較では、両エンジンとも学習前後で「同等」との評価が半数を占め、残る半数では「向上」と「劣化」が拮抗した。この時期は自動評価スコア、内容伝達レベルとも現状維持状態の期間にあたるが、第 3.3.2.項に示した個々の文単位の自動評価スコアの向上/低下カウント結果によって、この期間も全体の 4 分の 3 の機械翻訳文が学習の都度変化しており、それに伴う各文の自動評価スコアの変動が向上と低下とで拮抗した結果、全体平均スコアが毎回微動にとどまっていたということが判明している。本評価結果からは、この期間の翻訳品質の現状維持の理由が、より正確には「全体の

半数の文では学習前後で機械翻訳品質が実質的には変化しておらず、残り半数の文においては、自動評価スコアが示すとおり品質向上と劣化とが拮抗していたため」であることが示された。

続く第 10 回学習後と第 15 回学習後との比較では、全体の 40%強が「向上」、同じく 40%強が「同等」、残る約 10%が「劣化」と、前回に比べて「向上」と「劣化」のバランスが前者に大きく偏った。これは主に、第 12 回学習前に実施された「みんなの自動翻訳」の大規模アップデートの効果によるものである。自動評価スコアでは第 12 回学習後にスコアが顕著に向上し、第 15 回学習後もその水準が維持されたが、本評価結果もこれと整合している。

大規模アップデート以前は、一般／技術用語の訳語が学習前後でランダム的に変化する傾向が強く、不適訳語が突発的に採用されて品質を下げる文と、前回時に採用された不適訳語が適訳語に戻ったことで品質を上げる文とが拮抗する状況であったと考えられる。このため本評価でも、全体の半数の文が「向上」と「劣化」で拮抗する結果となっていた。これに対し、大規模アップデートを経た第 15 回後の評価では、学習データの内容に準拠した適訳語が順当に採用される確率が高まった結果、「不適訳語が突発的に採用されて品質を下げる文」の発生が抑えられ、かたや「不適訳語から適訳語に変化して品質を上げる文」は従前どおり発生したため、「向上」への大きな偏りが生じたと推察される。

この推察は、第 15 回学習後と第 20 回学習後との比較によってさらに裏付けられた。大規模アップデート以降の機械翻訳文同士の比較となる今回の比較では、100 文中汎用 NT で 71 文、特許 NT では 81 文と全体の 70～80%が「同等」と評価された。大規模アップデート以前の翻訳文同士の比較となる第 5 回後と第 10 回後の比較結果と比べても「同等」と評価された文数は明らかに増加しており、大規模アップデート以降、各文の翻訳文において適訳語が安定的に採用されるようになった結果、学習前後で品質変化が生じない文が大多数となったことが改めて示された。第 3.3.2.項に示した文単位の自動評価スコアの向上／低下カウント結果では、この期間も全体の 3 分の 2 程度の文は毎回変化し続けており、このため「中国審決コーパスデータには学習効果とほぼ同程度の副作用が存在する可能性がある」ことが懸念されたが、本評価結果により、少なくとも大規模アップデート以降の「みんなの自動翻訳」では、大多数の翻訳文は学習前後で実質的に翻訳品質は同等であり、変化の大部分は内容理解に影響しない表層的なものであることが改めて確認された。

以上、本評価結果を通じて、自動評価スコアや内容伝達レベルから看取された各学習段階での全体的な翻訳品質の推移が、個々の文単位ではどのような内訳になっていたかを確認した。各学習段階での評価結果は、自動評価スコアや内容伝達レベルから看取した状況、すなわち、無学習時から第 5 回後にかけては特許・審決用語の全体的な改善によりほぼ全て

の文で品質が改善し、第 5 回後から第 10 回後にかけては技術／一般用語の不安定な訳語変化によって品質が向上する文と劣化する文とが同等に発生し、第 10 回後から第 15 回後にかけては大規模アップデートの効果で不安定な訳語変化が抑制されたことでこのバランスが向上に偏り、第 15 回後から第 20 回後にかけても引き続き訳語が安定して各文とも学習前後での品質変化が生じなくなった、という推移と整合するものであり、これらの推論の妥当性が裏付けられた。

4.5.3. 両エンジン間の機械翻訳文の相対評価

前項では翻訳エンジン（汎用 NT、特許 NT）別に学習前後の機械翻訳文の優劣を判定したが、本観点では、二種のエンジンの機械翻訳文同士を比較して、その優劣を判定した。すなわち、前観点が「(同じエンジンの) 学習前後の機械翻訳文のどちらが優れているか」の評価であるのに対し、本観点は「二種の翻訳エンジンの（学習後の）機械翻訳文のどちらが優れているか」の評価となる。

評価は、一方を「優」、他方を「劣」とするか、双方を「同等」とするかのいずれかとした。内容伝達レベル自体に優劣がある場合はそれと一致した評価となるが、内容伝達レベルが同値であっても、品質に有意な差異がある場合は優劣がつけられた。各回の集計結果を下表に示す。

表 4.5.3. 各回学習前後の両エンジン間の相対評価結果

学習前	汎用 NT が優れる	両エンジン同等	特許 NT が優れる
第 5 回学習後	43	37	20
第 10 回学習後	28	50	22
第 15 回学習後	8	79	13
第 20 回学習後	12	79	14

自動評価では、中国審決コーパスデータ全件学習後の二種類のエンジン間にスコア上の優劣はなく、両エンジンのどちらを使用しても同等の翻訳品質が得られることが示された（⇒3.3.1.）。また、両エンジンは基礎学習データが一部異なるが、両エンジンの各分野の技術用語の翻訳品質には顕著な差異は存在しないとも結論した（⇒3.3.3.3.）。本評価結果において、これらの考察内容の妥当性を検証するとともに、個々の文における詳細な状況を把握する。

上表に示した評価結果からは、第 5 回学習後こそ優劣が汎用 NT に偏ったものの、以降は各回とも 100 文トータルでは両エンジンの優劣は常時ほぼ拮抗していたことが見て取れる。特に大規模アップデートを経た第 15 回学習後以降は「両エンジン同等」という評価が大勢を占めるようになり、100 文全体としてのみならず、個々の文単位でも両エンジンの翻訳品質の均質化が進んだことがわかる。最終学習後は全 100 文中のほぼ 8 割に相当する 79 文が「どちらの翻訳エンジンでも同等の翻訳品質が得られる」と判定されており、自動評価結果から導出した上記結論が裏付けられた。

4.5.4. 実例を用いた機械翻訳文の変遷の分析

前項まで、内容伝達レベル、学習前後の相対評価、そして翻訳エンジン間の相対評価の各観点について、人手評価の集計結果を示してきた。本項では、実際に人手評価の対象となった実例をピックアップし、各回の機械翻訳文を示しながら、ここまで述べてきた考察内容を確認していく。

4.5.4.1. 実例 4-1：文番号#1494、汎用 NT の分析

最初の実例として、文番号#1494 の汎用 NT の機械翻訳結果を分析する。まずは中国審決コーパスデータを学習させる前の、デフォルト（無学習）状態の汎用 NT による機械翻訳結果を以下に示す。

【文番号：#1494、汎用 NT】

原文	证据 2 公开一种抽屉滑轨组，其没有公开在抽屉 40 以及抽屉容室 50 的两侧水平设置滑轨的内容，没有公开上述区别特征，也没有给出相关启示。	
正解訳文	証拠 2 は引き出しとスライドレールのユニットを開示しているが、引き出し 40 と、引き出し収納室 50 の両側にスライドレールを水平に設置するという内容を開示しておらず、つまり上記相違点を開示していないため、関連の示唆を与えていない。	
無学習 内容伝達 Lv：2	証拠 2 は引き出しスライドレール組を公開する、それは引き出し 40 及び引き出し収納室 50 の両側水平にスライドレールを設置する内容を公開しない、前記区別特徴を公開しない、関連の啓発も提供しない。2。	
	—	BLEU：0.1974（－） RIBES：0.7955（－）

上掲した無学習時の機械翻訳結果に対しては、人手評価者により多数の不備が指摘された。冒頭から順に述べると「引き出しスライドレール組」という技術用語の誤訳（これでは正解訳文の「引き出しとスライドレールのユニット」とは伝わらない）、「公開する」、「公開しない」という不自然な日本語用法（終止形が不適）、「両側」の係り受けの誤り（助詞が抜けており係り受けが不明）、「区別特徴」と「啓発」という審決用語の誤訳（正解訳文の「相違点」「示唆」が適訳）、そして文末における不要な「2。」の出力が挙げられる。これら多数の不備により、内容伝達レベルもレベル 2「いくつかの重要情報は正確に伝達されている（20%～）」という低評価が下された。汎用 NT の無学習時における内容伝達レベルの平均は 2.46 であり、無学習の時点では、他の多くの文も本例と大差ない翻訳品質であったということになる。

これに対し、第1回人手評価時、すなわち第5回学習後の機械翻訳結果は以下のように変化した。

第5回学習後 内容伝達Lv:4	証拠2は引き出しスライドレール群を開示しており,引き出し40及び引き出し収容室50の両側にスライドレールが水平に設置される内容を開示しておらず,上記相違点を開示しておらず,関連する示唆も与えていない。2。	
	向上	BLEU:0.4673 (+0.2699) RIBES:0.8801 (+0.0846)

前回からの改善点を挙げると、不自然な日本語用法として指摘されていた終止形の誤用(「公開する」、「公開しない、」)が解消し、日本語文としての違和感がなくなった。「両側」の係り受けも正しくなり、審決用語の誤訳(「区別特徴」「啓発」)も、それぞれ「相違点」「示唆」と正解訳文どおりの適訳語に改善された。

一方、「引き出しスライドレール群」という技術用語の誤訳と文末の不要な「2。」の出力に関しては改善されなかった。また「開示しており」という表現についても、前回時の「公開する、」のような日本語用法上の誤りは解消したが、本来は正解訳文「開示しているが」のように逆接とするのが日本語として自然であり、引き続き軽微な日本語表現の不備と判定された。内容伝達レベルにおいては、これらのエラーのうち「引き出しスライドレール群」という技術用語の誤訳が、全くの間違いではないにせよ「群」では「ユニット」であることが正しく伝わらず、無学習時の「引き出しスライドレール組」と同等かそれ以下と判断された結果、レベル4に抑えられた。ただし、この点を除けば原文の内容はほぼ正確に把握でき、無学習時と比べて大幅な品質向上を遂げている。なお、第5回学習後の汎用NTの内容伝達レベルの全体平均は4.43(汎用NT)とこちらも約2レベル向上しており、多くの文で本例と同程度の品質改善がなされたと見なせる。

続いて、第2回人手評価時(第10回学習後)の機械翻訳文を示す。

第10回学習後 内容伝達Lv:3	証拠2は引き出しスライドレール群を開示しており,引き出し40及び引き出し収容室50の両側にスライドレールの内容を水平に設置するという内容を開示しておらず,上記相違点についても開示しておらず,関連する示唆も与えていない。2	
	劣化	BLEU:0.5075 (+0.0402) RIBES:0.8727 (-0.0074)

第10回後の機械翻訳文は、一箇所を除けば前回時とほぼ同等の内容といえるが、その一箇所の変化によって内容伝達レベルは3に下げられた。具体的には、文中赤字で示した

「内容を」の箇所である。機械翻訳文を見ると、正しい位置にも「内容を」は別途出力されており、つまり同じ語が重複して訳出されたことになる。このため本件は、原文に存在しない要素が機械翻訳文に混入する「湧き出しエラー」と判定された。内容理解への悪影響も大きく、重大なエラーである。

上記湧き出しエラー（「内容を」の重複）は、前掲した無学習時と第5回学習後の機械翻訳文では発生していない。さらに言えば、人手評価の対象外である第1回～第4回及び第6回～第9回学習後の機械翻訳文、そして特許NTの各回の機械翻訳文でも発生しておらず、汎用NTの第10回学習後のみで突発的に発生したエラーである。当時の「みんなの自動翻訳」では、本例のように原文に存在しない要素が翻訳文に混入する湧き出しエラーや、その逆の、原文中の要素が翻訳文から消失する欠落エラーの突発的な発生が散見された。第3.4.項に示した実例定点観測でも、例えば実例3-3では第2回～第4回後にかけて「重合」の湧き出しや、第1回及び第3回後における「解乳化ステップ」の欠落が発生している。

これらの湧き出しエラー、欠落エラーは、ある学習回のみで突如発生し、後続の学習回では速やかに解消する傾向が強く、また同一学習回であっても一方のエンジンのみが発生することが多い。こうした傾向は、技術／一般用語における突発的な不適訳語の採用と類似しており、どちらも根本的にはエンジンのロジックそのものに起因するものと考えられる。

なお、前回時に適訳に改善された各審決用語（「相違点」と「示唆」）は、今回も同じ適訳語が採用されている。初回学習以降、人手評価対象外の各回でもこの適訳語が常時安定して採用され続けており、定点観測した各実例で見られた「特許・審決用語の初期学習時から常時安定する」傾向が本例にも表れている。

第10回学習後の機械翻訳文は、「内容を」の湧き出しエラーにより品質が低下したが、突発性の強いこのエラーを除けば、他の変化箇所は「相違点を開示しておらず」から「相違点についても開示しておらず」への、一般用語間の同義の範囲内の変化のみであった。したがって第10回後の機械翻訳文も、この時点の自動評価スコアや内容伝達レベルが示した全体傾向、すなわち「実質的な品質は現状維持」という傾向と、基本的には整合しているといえる。

次に、第15回学習後の機械翻訳結果を示す。前回（第10回）の評価後、第12回学習前に「みんなの自動翻訳」の大規模アップデートが実施されており、今回の機械翻訳文からは大規模アップデート後のエンジンによる機械翻訳文となる。

第 15 回学習後 内容伝達 Lv : 4	証拠 2 は引き出しスライドレールセットを開示しており,それは引き出し 40 及び引き出し収容室 50 の両側にスライドレールを水平に設置するという内容を開示しておらず,上記相違点を開示しておらず,関連する示唆も与えていない。	
	向上	BLEU : 0.5663 (+0.0588) RIBES : 0.9086 (+0.0359)

上記翻訳文で特記すべきは、過去各回で常時発生していた、文末の不要な「2」の出力が解消されたことである。このエラーは原文の「両」という文字が原因とみられ、この文字を含む文ではきわめて高確率で発生する。本例においても、無学習時から第 11 回学習後にかけて、特許 NT の第 3 回、第 6 回学習後を除く全ての回でこのエラーが発生している。

本エラーの発生メカニズムについては後述の第 4.5.5.8.項「エラーのカテゴリ別カウント」の「数値エラー」の項で詳述するが、「両」を含む全ての文で無学習の時点から毎回ほぼ確実に発生していたことから、明らかにエンジンのロジックが関与するエラーである。第 12 回学習前の大規模アップデート以降、本例を始め全ての文において本エラーが一斉に解消され、以後全く再発しなくなった。これは、大規模アップデート時に本エラーへの抜本的な対策が講じられたためと考えられる。

大規模アップデート時の品質向上について、これまでその主な理由を「一般／技術用語のランダム的な変化が抑制され、学習データに準拠した適訳語が安定して採用される確率が高まったため」としてきたが、本例により、エンジンのロジックに起因するエラーがアップデートにより除去されたことも品質改善の一因であることが判明した。

なお、今回の翻訳文では前回突発的に発生していた「内容を」の湧き出しエラーも解消している。厳密には、本エラーは発生直後の回（第 11 回学習後）で既に解消しており、以降も再発していない。このランダム性、突発性が「みんなの自動翻訳」の湧き出しエラーの特徴である。

前々回、前回と技術誤訳判定された「引き出しスライドレール群」は、今回「引き出しスライドレールセット」と末尾部分が変わったが、依然として正しい意味で理解することとは難しく、引き続き誤訳と見なされた。

第 15 回学習後の翻訳文の内容伝達レベルは 4 であり、前回時より 1 レベル向上した。前回突発的に発生した「内容を」の湧き出しエラーが解消したことが大きいですが、文末の不

要な「2」の出力も解消されている。こうしたエンジンのロジックに起因する各種エラーが一斉に解消されたことも、技術／一般用語において不適訳語の採用が抑制されたことと同様、大規模アップデートの主要な品質改善効果であったことがうかがえる。本例に見られる改善傾向は、自動評価スコア及び内容伝達レベルから看取される全体傾向と引き続き合致している。

最後に、第 20 回学習後の機械翻訳文を示す。

第 20 回学習後 内容伝達 Lv : 4	証拠 2 は引き出しスライドレールセットを開示しており,それは引き出し 40 及び引き出し 収納室 50 の両側にスライドレールを水平に設置するという内容を開示しておらず,上記相違点を開示しておらず,関連する示唆も与えていない。	
	向上	BLEU : 0.5963 (+0.0300) RIBES : 0.9192 (+0.0106)

第 20 回学習後の翻訳文は前回時（第 15 回後）とほぼ同じで、唯一の変化は「厩容室」の訳語が「引き出し収容室」から「引き出し収納室」に変わったことのみである。正解訳文と同じ訳語となったため自動評価スコアは向上しているが、同義の範囲であり、懸案の「引き出しスライドレールセット」の訳語は改善していないため、内容伝達レベルは 4 のままであった。

本例の各評価回の機械翻訳文の変遷からは、無学習時の「みんなの自動翻訳」ではほとんど対応できていなかった特許・審決用語が中国審決コーパスデータの学習の結果、早期から適訳語に訳されるようになり大幅な品質改善がなされたこと、ただし大規模アップデート以前の旧「みんなの自動翻訳」では、文末における不要な「2」の出力や、第 10 回後の「**内容**」の重複など、エンジンのロジックに起因するエラーが頻繁に発生していたこと、そして大規模アップデート以降はこうしたエンジン由来のエラーが解消されるとともに、訳語の不安定な変化が抑制されるようになり、学習データの内容を的確に反映した順当な翻訳文が安定的に出力されるようになったことなど、自動評価や人手評価の全体平均スコアが端的に示していた状況が具体的に確認できた。

4.5.4.2. 実例 4-2 : 文番号#704、汎用 NT の分析

次の事例は文番号#704、汎用 NT による機械翻訳結果である。まずは原文、正解訳文と無学習時点での機械翻訳結果を示す。

【文番号：#704、汎用 NT】

原文	⑤在新穎性的判断中，对比文件公开的技术内容除文字记载的内容外还包括对于本领域技术人员来说隐含公开的技术内容。	
正解訳文	⑤新規性を判断する場合に、引用文献で開示されている技術内容は、文字によって記載された内容以外に、さらに当業者にとって実質的に開示されている技術内容も含む。	
無学習 内容伝達 Lv : 3	⑤新規性の判断において、文書公開の技術内容は、文字記載の内容以外にも、本分野の技術者にとって暗黙的に公開されている技術内容を含んでいる。	
	—	BLEU : 0.1949 (—) RIBES : 0.6739 (—)

無学習時の機械翻訳文の内容伝達レベルは3と判定された。重大なエラーは前半の「文書公開の」の箇所であり、本来は正解訳文のように「引用文献で開示されている」とすべきところ、原文の「对比」に相当する「引用」が訳出されておらず、重要な情報が欠落している。さらに後半の「公開の」の部分も、ひとまず意味は通るが日本語としてやや乱暴であり、日本語表現に改善の余地がある。

ただし、内容理解を妨げるレベルのエラーはこの一箇所のみである。それ以外は、「当業者」「実質的に開示」に相当する審決用語がそれぞれ「本分野の技術者」「暗黙的に公開」と直訳調で訳され、誤訳と判定されたが、意味はほぼ正確に理解でき、内容伝達の観点からは致命的ではない。これらの状況を総合して、内容伝達レベルは3と判定された。

続いて第5回学習後の機械翻訳結果を示す。

第5回学習後 内容伝達 Lv : 4	⑤新規性の判断において、引用文献に開示された技術内容は文字に記載されている内容に加えて、当業者に実質的に開示されている技術内容を含む。	
	向上	BLEU : 0.4002 (+0.2053) RIBES : 0.8276 (+0.1537)

大きな改善点として、無学習時にエラーと判定された三箇所、「文書公開の」「本分野の技術者」「暗黙的に公開」がいずれも「引用文献に開示された」「当業者」「実質的に開

示」と正解訳文どおりの適訳語に改善された。つまり前回時にエラーと判定された全ての箇所が正されたことになる。これら三箇所はいずれも審決用語・表現に関する箇所であり、他の各事例と同様、全て学習初期の時点で適訳語に改善された。

ただし今回の機械翻訳文には、前回時は妥当に訳されていたのに劣化してしまった箇所も存在する。具体的には「文字に記載されている内容」という箇所であり、これでは本来の「文字で記載された内容」というニュアンスが伝わらない。この箇所は前回時は「文字記載の内容」と訳されており、舌足らずながら正しい意味に解釈できる余地はあった。このエラーにより内容伝達レベルは今回も最高評価とはならず、レベル4にとどめられた。それでも前回時よりは1レベル向上しており、主に特許・審決用語の適訳語への改善によって、翻訳品質は確実に改善している。

続いて第10回学習後の機械翻訳結果を示す。

第10回学習後 内容伝達Lv:3	⑤新規性の判断において、引用文献に開示された技術内容は文字に記載された内容以外にも当業者に広く開示されている技術内容を含む。	
	劣化	BLEU: 0.3904 (-0.0098) RIBES: 0.7140 (-0.1136)

第4.5.1.項で示したとおり、第10回学習後の汎用NTの内容伝達レベルの全体平均値は前回(第5回後)から事実上の現状維持と見なされた。本例にもその傾向は顕れており、前回時の係り受けエラー「文字に記載された内容」は今回も改善されていない。かつ本例においては今回、前回時は「実質的に開示」と定訳で訳されていた箇所が「広く開示」と誤ったニュアンスで訳されてしまった。このため内容伝達レベルは1レベル下げられ、レベル3と判定された。

本調査分析では、「みんなの自動翻訳」の顕著な特徴として、特許・審決用語が学習初期から最適訳で安定することを指摘してきた。この傾向はきわめて堅固であり、この事例でも「引用文献」や「当業者」は今回も定訳語で安定している。このような中、頻出の審決表現である「実質的に開示」が今回「広く開示」と誤訳されてしまったことは、上記特徴と矛盾する³⁰。

だが、特許NTと併せた2エンジン、初回学習後(第1回)から最終回(第20回)学習後までの全40種の機械翻訳結果において、この箇所が「広く」と訳されたのはこの一

³⁰ ただし、第3.4.5.項で取り上げた実例3-5でもそれまで「登録日」と正しく訳されていた箇所が第7回学習後の1回のみ「発行日」と誤訳されており、特許・審決用語の突発的な誤訳も皆無ではない。

回のみであり、それ以外の 39 回は全て「実質的に」と訳されている。そもそも原文の「隠含」は「明記されていないが」というようなニュアンスであり、「広く」とは全く意味が異なる。このことから、本エラーは、学習データ中に「隠含」を「広く」と訳している事例が存在するためではなく、「隠含」に相当する語が何らかの理由で欠落した埋め合わせとして、中国審決コーパスデータ中で「開示」と組み合わせて使用されることが多い「広く」という語が出力された可能性がある。

第 10 回当時の「みんなの自動翻訳」では、本例のようにそれまで問題なく訳されていた語が突如欠落したり、原文中に存在しない語が突発的に湧き出したりするケースが散見された。本例も、エンジンのロジックに起因する突発的なエラーが審決用語において発生したケースと捉えられる。

続いて第 15 回学習後の機械翻訳結果を示す。

第 15 回学習後 内容伝達 Lv : 5	⑤新規性の判断において、引用文献に開示されている技術内容は文字で記載されている内容以外に、当業者にとって実質的に開示されている技術内容をさらに含む。	
	向上	BLEU : 0.5507 (+0.1603) RIBES : 0.9014 (+01874)

第 15 回後の機械翻訳文は、「みんなの自動翻訳」の大規模アップデート後のものとなる。前回時の重大な特許誤訳「広く開示」は、突発性と推測したとおり、今回は「実質的に開示」と正しい内容に復帰した。さらに、前々回、前回と連続でエラーとなっていた「文字に記載された内容」の箇所も、今回は「文字で記載されている内容」と正しい係り受けに改善されている。これにより機械翻訳文に一切の不備は無くなり、内容伝達レベルも最高位のレベル 5 が付された。

自動評価結果を分析した第 3 章では、大規模アップデートによる翻訳品質改善の主要因として、学習データ中の有力な訳語候補が順当に採用される確率が高められ、学習データの内容がよりの確に機械翻訳文に反映されるようになったことを挙げた。今回の機械翻訳文における「文字で記載されている内容」への改善にも、その効果がうかがえる。以下、全学習回における当該箇所の訳語の推移を示す。

表 4.5.4. 実例#704 における「文字で記載されている内容」の訳語変化状況

エンジン	汎用 NT	特許 NT
無学習	文字記載の内容	文字記載の内容
第 1 回	文字で記載された内容	文字で記載された内容
第 2 回	文字による記載内容	文字に記載された内容
第 3 回	文字で記載されている内容	文字で記載されている内容
第 4 回	文字で記載されている内容	文字による記載内容
第 5 回	文字に記載されている内容	文字による記載内容
第 6 回	文字で記載された内容	文字で記載された内容
第 7 回	文字で記載されている内容	文字で記載されている内容
第 8 回	文字に記載されている内容	文字で記載されている内容
第 9 回	文字で記載されている内容	文字で記載された内容
第 10 回	文字に記載された内容	文字に記載された内容
第 11 回	文字に記載されている内容	文字で記載されている内容
第 12 回	文字で記載されている内容	文字で記載されている内容
第 13 回	文字で記載されている内容	文字で記載されている内容
第 14 回	文字で記載されている内容	文字で記載されている内容
第 15 回	文字で記載されている内容	文字で記載されている内容
第 16 回	文字で記載されている内容	文字で記載されている内容
第 17 回	文字で記載されている内容	文字で記載された内容
第 18 回	文字で記載されている内容	文字で記載されている内容
第 19 回	文字で記載されている内容	文字で記載されている内容
第 20 回	文字で記載された内容	文字で記載されている内容

汎用 NT の人手評価回では第 5 回後、第 10 回後と「文字に記載」という不適訳が続けて採用されたが、上表に示したとおり、この期間も人手評価対象外の回では「文字で記載」又は「文字による記載」と正しい解釈で翻訳されていることが多く、その合間で散発的に「文字に記載」という不適訳が採用されるというのが実際の状況であった。つまり、複数の候補から毎回ランダム的に訳語が選定され、突発的に不適訳語が採用されてしまうことがあるという、大規模アップデート以前の「みんなの自動翻訳」における一般用語・表現の傾向どおりの状況であった。

これに対し、大規模アップデートを経た第 12 回学習後以降は、両エンジン全ての回で「文字で記載」という正しい係り受けが安定的に採用されるようになり、「文字に記載」という誤訳は一切採用されなくなった。つまり本例も、大規模アップデートを境に、それ

までランダム的に採用されていた複数の訳語候補のうち、学習データ中で最も使用頻度の高い訳語が毎回安定して採用されるようになった典型例といえる³¹。

最後に、第 20 回学習後の機械翻訳結果を示す。

第 20 回学習後 内容伝達 Lv : 5	⑤新規性の判断において、引用文献に開示されている技術内容は文字で記載されている内容以外に、当業者にとって実質的に開示されている技術内容をさらに含む。	
	同等	BLEU : 0.5507 (±0) RIBES : 0.9014 (±0)

第 20 回後の機械翻訳文は第 15 回後のものと完全に同一であった。大規模アップデート以降、訳語の不安定な変化が抑制されたことで各機械翻訳文において学習前後での変化が局所化・小規模化し、前回時から翻訳文が「不変」となる文も顕著に増加したが、本例はその典型例といえる。また、自動評価や内容伝達レベルが示した、大規模アップデート時に向上した品質水準が以降も常時維持されたという状況とも整合している。

以上、本例の機械翻訳文の変遷からは、表 4.5.4. に示した「文字で記載されている内容」の箇所に採用された訳語の変化状況に代表されるように、大規模アップデートの効果によって翻訳文の各所において訳語が適訳語で安定化し、突発的な不適訳の採用が抑制されるようになった結果、翻訳品質が改善され、そのまま安定するという状況が具体的に見て取れた。

³¹ 「中国審決コーパスデータ」では「文字で記載」という訳し方が圧倒的多数である。

4.5.4.3. 実例 4-3 : 文番号 #830、汎用 NT の分析

本例は、各回の機械翻訳結果を一覧で示して解説する。

【文番号 : #830、汎用 NT】

原文	2、権利要求 2 在権利要求 1 基础上进一步限定的技术特征已被证据 1 完全公开, 该权利要求 2 与证据 1 相比不具备新颖性和创造性。	
正解訳文	2、請求項 2 が請求項 1 を基礎としてさらに限定した技術的特徴は既に証拠 1 により完全に開示され, 該請求項 2 は証拠 1 と比較して新規性及び創造性を有しない。	
無学習 内容伝達 Lv : 2	2.権利要求 2 は,権利要求 1 に基づいてさらに限定された技術的特徴が, 証明 1 完全に公開されており,この権利要求 2 は,証拠 1 に比べて,新規性や創造性を持たない。	
	—	BLEU : 0.1050 (—) RIBES : 0.6845 (—)
第 5 回学習後 内容伝達 Lv : 3	2、請求項 2 は請求項 1 を基礎としてさらに限定している技術的特徴は証拠 1 に完全に開示されており,該請求項 2 は証拠 1 と比較して新規性及び創造性を有さない。	
	向上	BLEU : 0.6787 (+0.5737) RIBES : 0.9536 (+0.2691)
第 10 回学習後 内容伝達 Lv : 4	2、請求項 2 は請求項 1 に基づいてさらに限定されている技術的特徴が証拠 1 に完全に開示されており,該請求項 2 は証拠 1 と比較して新規性及び創造性を有さない。	
	向上	BLEU : 0.5434 (-0.1353) RIBES : 0.9162 (-0.0374)
第 15 回学習後 内容伝達 Lv : 5	2、請求項 2 が請求項 1 を基礎としてさらに限定している技術的特徴は証拠 1 に完全に開示されており,該請求項 2 は証拠 1 に対して新規性及び創造性を有さない。	
	向上	BLEU : 0.6388 (+0.0954) RIBES : 0.9328 (+0.0166)
第 20 回学習後 内容伝達 Lv : 5	2、請求項 2 が請求項 1 を基礎としてさらに限定している技術的特徴は証拠 1 に完全に開示されており,該請求項 2 は証拠 1 に対して新規性及び創造性を有さない。	
	同等	BLEU : 0.6388 (±0) RIBES : 0.9328 (±0)

本例の無学習時の翻訳文は、「請求項」と訳すべき「権利要求」という語を「権利要求」と直訳調で訳している点、そして文中に二度出現する「証拠 1」の 1 回目が「証明 1」となっており、かつ直後の助詞が欠落して係り受けが解釈不能となっている点において、大きく品質を損ねている。また、本文前半の主旨は「請求項 2 が請求項 1 をベースにこれをさらに限定した技術的特徴は、証拠 1 に開示されている」というものだが、無学習時のこの箇所は助詞が「請求項 2 は、請求項 1 に基づいてこれをさらに限定された技術的

特徴が,証拠1に…」と逆になっており、誤訳とまではいえないが日本語文として不自然で読みにくい。これらの不備により、内容伝達レベルはレベル2と低評価が下された。

第5回時学習後の翻訳文では、「権利要求」「証明1」といった特許用語の誤訳が解消し、以降各回とも安定して「請求項」「証拠1」と適訳語で訳されている。つまり本例においても、特許・審決用語が学習初期から最適訳で安定するという傾向は肯定された。

ただし第5回後の翻訳文は、前半部分の助詞が「請求項2は請求項1を基礎としてこれをさらに限定している技術的特徴は証拠1に…」と、文法的に成立しないレベルの係り受けエラーに悪化しており、このため内容伝達レベルは前回比1レベル向上のレベル3にとどまっている。

第10回学習後の翻訳文では、この箇所の助詞は「請求項2は請求項1を基礎としてこれをさらに限定している技術的特徴が証拠1に…」と無学習時の形に戻った。前回時よりは良化したものの日本語としては不自然であり、このため内容伝達レベルは満点にはならず、前回から1レベル向上のレベル4が付された。

大規模アップデートを経た第15回学習後の翻訳文では、ようやくこの箇所の助詞が正解訳文どおりの「請求項2が請求項1を基礎としてさらに限定している技術的特徴は証拠1に…」と改善された。他の箇所にも不備はなく、内容伝達レベルは最高位のレベル5となった。続く第20回学習後も、第15回後と完全に同一の翻訳文が出力された。

本例前半部分の助詞選択は、人手評価対象外の各回を見ると、大規模アップデート以前にも正解訳文どおりに訳されている回も数回ある（汎用NTの第1回、4回、6回、11回。特許NTでは皆無）。とはいえ、大規模アップデートを経た第12回学習後以降は、特許NTを含め、全ての回の翻訳文で正解訳文どおりの最適の助詞が安定して採用されており、アップデートを境に訳語選定の精度が明らかに改善していることがわかる。

なお、ここまで本項で取り上げた三件の実例では全て汎用NTの翻訳文を使用したか、特許NTの翻訳文の傾向も各例おおむね同じである。参考に、本例の特許NTの出力結果を以下に示す。

【文番号：#830、特許 NT】

無学習 内容伝達 Lv：2	2、請求項 2 に請求項 1 に基づいてさらに定義された技術的特徴は既に証拠 1 によって完全に開示され、該特許請求の範囲 2 及び証拠 1 は新規性及び創造性を有するものではない。
	－ BLEU：0.1050 (－) RIBES：0.6845 (－)
第 5 回学習後 内容伝達 Lv：3	2、請求項 2 は請求項 1 を基礎としてさらに限定している技術的特徴は証拠 1 により完全に開示されており、該請求項 2 は証拠 1 と比較して新規性及び創造性を有さない。
	向上 BLEU：0.6787 (+0.5737) RIBES：0.9536 (+0.2691)
第 10 回学習後 内容伝達 Lv：4	2、請求項 2 は請求項 1 を基礎としてさらに限定された技術的特徴が証拠 1 によって完全に開示されており、該請求項 2 と証拠 1 とを比較して新規性及び創造性を有さない。
	同等 BLEU：0.5434 (-0.1353) RIBES：0.9162 (-0.0374)
第 15 回学習後 内容伝達 Lv：5	2、請求項 2 が請求項 1 を基礎としてさらに限定している技術的特徴は証拠 1 に完全に開示されており、該請求項 2 は証拠 1 に対して新規性及び創造性を有さない。
	向上 BLEU：0.6388 (+0.0954) RIBES：0.9328 (+0.0166)
第 20 回学習後 内容伝達 Lv：5	2、請求項 2 が請求項 1 を基礎としてさらに限定している技術的特徴は証拠 1 に完全に開示されており、該請求項 2 は証拠 1 に対して新規性及び創造性を有さない。
	同等 BLEU：0.6388 (±0) RIBES：0.9328 (±0)

特許 NT では、無学習時において特許用語の翻訳品質が汎用 NT に比べてやや優れており、この点が汎用 NT との主な品質差となっていた。本例でも、汎用 NT では無学習時に「権利要求」と誤訳されていた「権利要求」が、特許 NT では「請求項」と正しく訳されている。ただし、後半では同じ「権利要求」が「特許請求の範囲」と訳ゆれしている。

また、汎用 NT で「証明 1 完全に」と大きく誤訳されていた箇所が特許 NT では「証拠 1 によって」と適切に訳されている。ただし、これは特許用語への対応の差というよりは、「みんなの自動翻訳」における訳語の不安定性によるものと考えらるべきである。こうした突発的な不適訳語の採用は特許 NT の翻訳文でも同等に発生しており、本例も後半で「特許請求の範囲 2 及び証拠 1」と、並列関係ではない要素が並列に扱われ、「証拠 1」が「新規性及び創造性を有さない」という意味に誤訳されている。また、汎用 NT では無学習時から一貫して「限定された」と問題なく訳されていた箇所が「定義された」とニュアンスの逸れた語で訳されている。これらを総合して、特許 NT の無学習時の内容伝達レベルは、汎用 NT と同じレベル 2 が付された。

続く第 5 回学習後の翻訳文は汎用 NT と実質的に同内容であり、違いは汎用 NT では「証拠 1 により完全に開示」であった箇所が「証拠 1 に完全に開示」となっていることのみである（両者は文法的な意味合いは異なるが、本例の文脈ではどちらでも正しい）。汎用 NT 同様、各特許用語は全て適訳語に改善され、その結果「請求項」と「特許請求の範囲」の訳ゆれも解消した。一方、「請求項 2 は…技術的特徴は証拠 1 に開示されており」という不適な助詞の使用に関しても同様に発生しており、内容伝達レベルも引き続き汎用 NT と同等のスコア 3 が付された。

この第 5 回学習後の結果を見る限り、無学習時に特許 NT のアドバンテージとなっていた特許用語（の一部）への対応度の差は、この時点で既に中国審決コーパスデータの学習効果によって完全に上書きされ、汎用 NT との間に実質的な品質差を生まなくなっている。基礎学習データの違いによる技術用語の翻訳品質にも両システム間に大きな差異はないと見られ（⇒3.3.3.3.）、そうなると、初期学習以降の両エンジンの優劣はほぼ「みんなの自動翻訳」の不安定性による偶然の誤訳や突発的なエラーの発生の有無に限られてくる。こうしたエラーは、そのランダム的な性質により一方のエンジンに大きく偏ることは考えにくく、自動評価スコアや内容伝達レベルの集計値もこれを肯定する結果となっている。よって中国審決コーパスデータを学習させる前提であれば、汎用 NT と特許 NT は、どちらを使用してもほぼ同等の品質の機械翻訳結果が得られるといえる³²。

続く第 10 回学習後の翻訳文でも、助詞の不適箇所は汎用 NT と同じく「請求項 2 は…技術的特徴が証拠 1 に開示されており」と変化しており、両エンジンの翻訳結果は同期している。内容伝達レベルも汎用 NT と同じく 1 レベル向上のスコア 4 が付された。

そして大規模アップデートを経た第 15 回学習後では、翻訳文自体が汎用 NT と完全一致した。アップデート以降、訳語のランダム性が抑制され、学習データの内容がよりの確に機械翻訳文に反映されるようになったこと、そしてその結果、同じ学習データを学習している両エンジン間においても訳語の均質化が進んだことを述べてきたが、本例もその証左といえる。翻訳文自体が同一であるため、内容伝達レベルも当然ながら汎用 NT と同じレベル 5 が付されている。

さらに第 20 回学習後の機械翻訳文も、第 15 回後と完全に同一のものが出力された。つまり大規模アップデート以降の第 15 回後と第 20 回後は、双方のエンジンで共通の翻訳文が連続して採用されたことになる。

³² 最終的な結論はその後の分析内容も加味して第 5 章 5.8.項で詳述する。

上記のとおり、第 5 回後、第 10 回後、第 15 回後、第 20 回後と、初回学習以降の内容伝達レベルは各回とも両エンジンで同等であった。実際の機械翻訳文を見ても、「特許用語への対応度の差」による特許 NT の優位性は中国審決コーパスデータの学習により早々に失われ、以降の両エンジンの差異は事実上「みんなの自動翻訳」の訳語のランダム性による誤差にすぎなくなった。さらに、このランダム性が抑制された大規模アップデート以降の第 15 回と第 20 回では、両エンジンの出力結果自体が完全に一致した³³。これらの結果は、第 4.5.3.項で述べた、両エンジン間の優劣は学習初期においては一定規模の文集合のトータルで同等であったが、その後さらに均質化が進み、最終的には個々の文単位でもおおむね同等の翻訳品質が得られるようになったという考察と整合している。

4.5.4.4. 実例 4-4：最終学習後に内容伝達レベルが最も低かった実例の分析

第 4.5.1.項に示したとおり、第 20 回学習後は全体の 97%の文が内容伝達レベル 4~5 を付されており、レベル 3 を付された 6 文（両エンジン 3 文ずつ）が「最終学習後に最も内容伝達レベルが低かった文」となる。具体的には汎用 NT では文番号#462、#1521、#1890 の 3 文、特許 NT では#462、#1521、#1743 の 3 文であり、#462 と#1521 は両エンジンでレベル 3 となっている。本項では、これらレベル 3 が付された各機械翻訳文の内容を分析し、その低評価の原因について考察する。

まずは両エンジンでレベル 3 となった文番号#462 の第 20 回学習後の機械翻訳文を以下に示す。

【文番号：#462】

原文	对于前置审查意见中的主要观点, 合议组认为, 只有能够从对比文件的附图中直接地、毫无疑义确定的技术特征才属于公开的内容, 由附图中推测的内容, 或者无文字说明、仅仅是从附图中测量得出的尺寸及其关系, 不应当作为对比文件已公开的内容。
正解訳文	前置審査意見の主な観点に対し、合議体の判断は以下のとおりである：引用文献の図面から直接的、かつ一義的に特定することができる技術的特徴だけが開示された内容に属し、図面から推測される内容、或いは文字による説明がなく、単に図面から導き出された寸法と関係だけである場合は、引用文献に開示された内容とすることはできない。

³³ 人手評価対象外の回では両者が相違する場合もあったが、局所的かつ表層的な違いであり実質的な品質差は生じていない。

【汎用 NT】 第 20 回学習後 内容伝達 Lv : 3	前置審査意見における主な観点について、合議体の判断は以下のとおりである:引用文献の図面から直接、且つ一義的に確定できる技術的特徴のみが開示されている内容であり、図面から推測された内容、又は文字による説明がなく、 図面から測定して得られた寸法及びその関係は 、引用文献に開示されている内容とするべきではない。	
	—	BLEU : 0.4854 (—) RIBES : 0.9113 (—)
【特許 NT】 第 20 回学習後 内容伝達 Lv : 3	前置審査意見における主な観点について、合議体の判断は以下のとおりである:引用文献の図面から直接、且つ一義的に確定できる技術的特徴のみが開示されている内容に属し、図面から推測される内容、又は文字による説明がなく、 図面から測定して得られた寸法及びその関係は 、引用文献に開示されている内容とするべきではない。	
	—	BLEU : 0.5202 (—) RIBES : 0.9233 (—)

両エンジンの機械翻訳文はほぼ同内容であり、レベル 3 が付される原因となった箇所も同一である。その箇所は具体的には、正解訳文で述べられているように「引用文献に文字で明記されておらず、添付図から見て取れる寸法やその関係性のみを根拠としている場合、引用文献が開示している内容とはみなされない」という主旨である。機械翻訳文は両エンジンともこの箇所を「**図面から測定して得られた寸法及びその関係は**」と訳しており、「そのみ根拠である場合は」という重要なニュアンスが表現できていない。このことが重大視され、他には特段の不備がないにもかかわらず、レベル 3 が付された。原文では「仅仅」がこのニュアンスを表しており、この表現自体は中国審決コーパスデータ中でも数千文で使用されている一般表現であるが、両エンジンとも同じ形で誤訳されていることから、本例に近い文脈での使用例は学習データ中に存在していない可能性が高い。

続いて、こちらも両エンジンともレベル 3 が付された文番号#1521 の最終学習後の機械翻訳文を分析する。

【文番号 : #1521】

原文	专利权人认为：(1) 关于专利法第 26 条第 4 款，权利要求 1 中的“不反应”是指“介在相与硬磁相或介在相与软磁相之间相互作用形成新的物质或者相，从而导致 软硬两项 距离变长的过程”，因此权利要求的限定是清楚的。
正解訳文	特許権者の主張は以下のとおりである：(1) 特許法第 26 条第 4 款について、請求項 1 の「反応しない」とは、「 介在相 と硬質磁性相又は軟磁性相との 間の相互作用 が新規の物質又は相を形成することによって、 硬軟両相 の距離が長くなる過程」を指しており、よって請求項の規

	定は明確である。
【汎用 NT】 第 20 回学習後 内容伝達 Lv : 3	特許権者の主張は以下のとおりである:(1)特許法第 26 条第 4 款について、請求項 1 における「反応しない」とは「前置相と硬磁性相又は前置相と軟磁性相との間の相互作用により新たな物質又は相を形成し、それにより硬さの二つの項目の距離が長くなる過程」を指すため、請求項の限定は明確である。
	— BLEU : 0.6104 (-) RIBES : 0.9222 (-)
【特許 NT】 第 20 回学習後 内容伝達 Lv : 3	実用新案権者の主張は以下のとおりである:(1)特許法第 26 条第 4 款について、請求項 1 における「反応しない」とは「中間相と硬磁性相又は中間相と軟磁性相との間に相互作用して新たな物質又は相を形成し、それにより硬と軟の二つの距離が長くなる過程」を指すため、請求項の限定は明確である。
	— BLEU : 0.5731 (-) RIBES : 0.9110 (-)

#1521 に関しては、両エンジンの翻訳文ともレベル 3 が付されているものの、その理由はそれぞれ異なる。汎用 NT の場合は、重要な技術用語である「介在相」が「前置相」と誤訳されていることと、「硬軟両相の」に相当する部分が「硬さの二つの項目」となっていて「軟」にあたる語が欠落しており、かつ「相」であることも明示されていないことが、ともに重大なエラーとして減点対象となった。一方、特許 NT では、「介在相」は「中間相」と同義の範疇の訳語で無難に訳されているが、「硬軟両相の」の部分は「硬と軟の二つ」となっており、こちらは「軟」は欠落していないが「相」のニュアンスがやはり明示されていない。特許 NT にはさらに「特許権者」が「実用新案権者」となる「專利」の訳し分け誤り（特許誤訳）、そして「(二種の相の間の相互作用が)とすべきところ「二種の相の間に相互作用して」と誤った係り受け解釈も発生しており、これらを総合してレベル 3 と判定された。こうしたエラーの多くは前回時（第 15 回後）には発生しておらず、「みんなの自動翻訳」特有の突発的な不適訳語の採用が複数箇所が発生した結果の低評価と見なせるが、両エンジンとも上手く訳せていない「硬軟両相の」の部分のみは、前例と同様、学習データ中に参考となる文例が存在しない可能性が高い。

続いて、一方のシステムのみでレベル 3 となった 2 文についても実例を確認する。まずは汎用 NT のみでレベル 3 となった文番号#1890 を示す。

【文番号：#1890】

原文	復審請求人认为：根据本申请的发明内容部分，本申请的目的是提供一种手段来有效地管理大规模的、 企业范围的 、与信息技术有关或与电信有关的 迁移 。	
正解訳文	復審請求人の主張は以下のとおりである：本願の発明の概要部分によると、本願の目的は大規模な、 全企業的な 、情報技術に関連するか又は電気通信に関連する マイグレーション を効果的に管理する解決手段を提供することである。	
【汎用 NT】 第 20 回学習後 内容伝達 Lv：3	復審請求人の主張は以下のとおりである：本願の発明の概要部分によれば、本願の目的は、情報技術に関連するか又は電気通信に関連する大規模な 移行 を効果的に管理する手段を提供することである。	
	—	BLEU：0.7264 (－) RIBES：0.9510 (－)
【特許 NT】 第 20 回学習後 内容伝達 Lv：4	審請求人の主張は以下のとおりである：本願の発明の概要部分に基づき、本願の目的は大規模な、 企業規模の 、情報技術に関連する又は電気通信に関連する 移行 を効果的に管理する手段を提供することである。	
	—	BLEU：0.7311 (－) RIBES：0.9510 (－)

本例の原文にある「**企业范围的**」は、あえて日本語にすれば正解訳文のように「**全企業的な**」というようなニュアンスと考えられるが、そもそも日本語ではこうした文脈でこの意味の語句を用いる慣習がなく、翻訳難易度が高い。このため学習初期の機械翻訳文では「**企業範囲**」など意味不明の訳語が採用されていたが、第 20 回学習後の特許 NT の翻訳文では上掲のとおり「**企業規模の**」と、かなりニュアンスに近い訳語に改善された。

これに対し、同じく第 20 回学習後の汎用 NT の機械翻訳文では、正解訳文の「**全企業的な**」に相当する語句が完全に欠落してしまっている。この重大な欠落に加え、「**マイグレーション**」と訳すべき技術用語を「**移行**」という不適訳語で訳している点も減点対象となり、総合でレベル 3 となった（特許 NT も同じく「**移行**」という不適訳語が採用されているためレベル 4 にとどまっている）。

汎用 NT で欠落した「**企业范围的**」という語は難訳語であることは確かだが、前回時（第 15 回後）の翻訳文では汎用 NT も「**企業規模で**」と妥当に訳出できており、第 20 回後の欠落は、エンジンのロジックに起因する突発的な欠落と見なせる。

最後に、特許 NT のみレベル 3 判定となった文番号#1743 の実例を示す。

【文番号：#1743】

原文	专利权人认为：(1) 证据 1 中具有臭氧吸出装置，从根源上断绝结晶状异物的产生，不存在本专利中的结晶状异物，是本专利首次发现的。
正解訳文	特許権者の主張は以下のとおりである：(1) 証拠 1 にはオゾン吸引装置があり，結晶状異物の発生を根本から断絶し，本特許における結晶状異物が存在せず，それは本特許が初めて発見したことである。
【汎用 NT】 第 20 回学習後 内容伝達 Lv：4	実用新案権者の主張は以下のとおりである：(1) 証拠 1 はオゾン吸引装置を有し，根本的に結晶状異物の発生を遮断し，本実用新案における結晶状異物が存在せず，本実用新案で初めて発見されたものである。
	— BLEU：0.5417 (－) RIBES：0.9182 (－)
【特許 NT】 第 20 回学習後 内容伝達 Lv：3	実用新案権者の主張は以下のとおりである：(1) 証拠 1 にはオゾン吸引装置があり，根本的に結晶性異物の生成を遮断しており，本実用新案における結晶性異物が存在しないことは，本実用新案で初めて発見されたことである。
	— BLEU：0.4764 (－) RIBES：0.8756 (－)

この文は正解訳文自体がやや難解な訳し方となっているが、原文の文意は、「証拠 1 の技術では結晶状異物はそもそも発生しないようになっており、(よって結晶状異物に対処する技術は) 本特許が初めて実現したものである」というものである。つまり「結晶状異物は証拠 1 の技術では発生しないが、本特許では発生する」と主張している。これに対し、上掲した特許 NT の機械訳文は、「本実用新案における結晶性異物が存在しないことは、本実用新案で初めて発見されたことである」となっており、本来の主張内容と逆の意味になってしまっている。重大な不備であり、その他の不備は「**專利**」の訳し分け誤りのみであるにもかかわらず、レベル 3 という低評価となった。なお汎用 NT はこの箇所は正解訳文どおりに無難に訳しており、不備は「**專利**」の訳し分け誤りのみであるためレベルは 4 が付されている。なお、この「**存在しないことは**」の箇所の誤訳は、第 20 回後こそ特許 NT のみでの発生だが、大規模アップデート以降の各回では両エンジンで頻発しており³⁴、そもそも機械訳訳にとって翻訳難易度が高い表現であることがうかがえる。

³⁴ 大規模アップデート以降 (第 12 回～第 20 回) の全 9 回中、特許 NT では 7 回 (第 12、13、15、17、18、19、20 回)、汎用 NT では 5 回 (第 14、15、16、17、18 回) 発生している。

以上、最終学習後にレベル 3 が付された各文のエラー状況を精査した。各文で低評価の主要因となっていたエラーには、翻訳難易度が高い用語や表現に起因する語句によるものと、突発的な不適訳語の採用によるものとに大別された。どちらも、仮に中国審決コーパスデータをさらに増強したとしても、完全に解消される性質のものではない。換言すれば、最も低評価であった機械翻訳文においても、中国審決コーパスデータの不備不足を示すようなエラー、つまり特許・審決用語や審決特有の表現等における重大なエラーは一切検出されなかった。このことから、本事業で作成した中国審決コーパスデータが質量ともに不足がないことが改めて確認された。

4.5.5. 「エラーのカテゴリ別カウント」結果

人手評価の一環として、機械翻訳の典型的なエラーを 10 種類カテゴリライズし、各文における発生の有無をカウントした³⁵。本項において各カテゴリのエラーのカウント結果を示すとともに、エラーの具体的な内容や、学習前後での改善の有無についてまとめる。

4.5.5.1. 「特許誤訳」のカウント結果

特許・審決特有の用語や表現に関する誤訳、不適訳は本カテゴリにカウントした。「請求項」「当業者」「口頭審理」などの特許・審決関連用語だけでなく、「容易に想到可能である」「(明細書に) サポートされていない」といった一般用語の審決特有の使い方や、定型文パターンの処理等に関する不備についても本カテゴリに含めた。各回のカウント結果は下表のとおりである。

表 4.5.5. 各回学習前後の「特許誤訳」該当文数カウント結果

エンジン	無学習	第 5 回学習	第 10 回学習	第 15 回学習	第 20 回学習
汎用 NT	91	5	5	15	9
	—	-86	0	+10	-6
特許 NT	90	21	13	16	16
	—	-69	-8	+3	±0

本事業で作成した中国審決コーパスデータの主用途は、機械翻訳システムの学習用データとして中国審決特有の用語や表現に適応させることである。その意味では、各種エラーの中でもこの特許誤訳の解消こそが本事業の主眼といえる。

上記カウント結果を見ると、無学習時、すなわちコーパス学習前のデフォルト状態では、人手評価対象文 100 文の約 90% の文で何らかの特許・審決用語の誤訳が発生していたことがわかる。無学習時の機械翻訳品質は内容伝達レベルがレベル 2 周辺と低評価であったが、こうした特許誤訳の多発がその主な原因であったことがうかがえる。

なお、各実例において基本的な特許用語への一定の対応が見られた特許 NT も、無学習時のカウント数は汎用 NT とほぼ同数となっている。これは、本カウントが語単位ではなく文単位であるためである。基本的な特許用語のいくつかが訳せていても、同じ文で審決特有の用語が誤訳されていればカウント上は同じ 1 文となる。この結果は、中国審決には基本的な特許用語のみで対応できる文がほとんど存在しないことを示している。

³⁵ 同一カテゴリのエラーを一文中で都度カウントする方式ではないため、カウントは最大で 100 (文) となる。

これに対し、コーパス学習後のカウントを見ると、第 5 回学習後の時点でエラー発生文数は汎用 NT では 5 文、特許 NT も 21 文と、いずれも急激に減少している。自動評価結果の分析では、各実例の機械翻訳文に見られる顕著な傾向として、特許・審決用語が学習初期から正解訳文どおりの適訳語で安定することを挙げたが、本カウント結果もこれを裏付けている。

以下、無学習時に特許誤訳と判定された実例と、コーパス学習後の改善状況について示す。

【文番号：#118、汎用 NT】

原文	对于上述 区别技术特征 ，由于 T7A2 等各启动子均为本领域常规使用的细菌或噬菌体启动子，因而 本领域技术人员 依据实际实验需求对构建体中的启动子进行替换并应用于蛋白表达系统中并不存在任何技术障碍。	
正解訳文	上記 相違点 について，T 7 A 2 等の各プロモーターはいずれも本分野で一般的に使用される細菌又はファージプロモーターであるため， 当業者 であれば実際の実験の要件に基づきコンストラクト中のプロモーターを置き換えかつタンパク質発現系に応用することにかかる技術的障害も存在しない。	
無学習 内容伝達 Lv：3	上述の 区别技術の特徴 に対して、T7A2 などの各プロモーターはすべて本分野で常用されている細菌またはバクテリオファージプロモーターであるため、 本分野の技術者 は実際の実験要求に基づいて、構築体中のプロモーターを入れ替えて、タンパク質発現システムに応用して、いかなる技術的な障害も存在しない。	
	—	BLEU：0.2295 (－) RIBES：0.8352 (－)
第 5 回学習後 内容伝達 Lv：5	上記 相違点 について、T7A2 等の各プロモーターはいずれも本分野で一般的に使用される細菌又はファージプロモーターであるため、 当業者 であれば、実際の実験要件に応じて構築物中のプロモーターを置換し、且つタンパク質発現系に応用することは、いかなる技術的障害も存在しない。	
	向上	BLEU：0.7395 (+0.5100) RIBES：0.9600 (+0.1248)

本例の原文には、主要な特許・審決用語として「**区别技术特征**」「**本领域技术人员**」の二語が含まれている。それぞれ正解訳文のとおり「**相違点**」「**当業者**」と訳すのが適切だ

が、無学習時の汎用 NT では「**区別技術の特徴**」「**本分野の技術者**」と直訳調の訳語となっており、どちらも特許誤訳と判定された³⁶。

このうち「**区別技術の特徴**」はそもそも理解不能な訳語であるが、「**本分野の技術者**」に関しては、原文の内容を理解する上ではこの訳語でも特に誤解は生じない。このような場合、内容伝達レベルの評価では重大視されないが、カテゴリ別エラーカウントでは、本語のように統一訳語が定まっている特許・審決用語の場合、原則としてその訳語でなければ特許誤訳と判定している。

このように複数の特許誤訳が発生していた無学習時に対し、中国審決コーパスデータを学習した第 5 回学習後の機械翻訳文では、両語とも「**相違点**」「**当業者**」と統一訳で訳されるようになり、その結果、文中の全ての特許誤訳が解消した。前掲のカウント結果でも第 5 回学習後には特許誤訳カウントが大幅に減少していたが、これは本例と同様、大多数の文において、無学習時に発生していた特許誤訳が全て解消したことを示している。

その後の各回の出力結果は以下のとおりである。

第 10 回学習後 内容伝達 Lv : 5	上記 相違点 について,T7A2 などの各プロモーターはいずれも本分野で一般的に用いられる細菌又はファージプロモーターであるため, 当業者 であれば実際の実験要件に応じて構築物中のプロモーターを置換し、且つタンパク質発現システムに応用することにかかる技術的障害も存在しない。	
	同等	BLEU : 0.6930 (-0.0465) RIBES : 0.9541 (-0.0059)
第 15 回学習後 内容伝達 Lv : 5	上記 相違点 について,T7A2 などの各プロモーターはいずれも本分野で一般的に使用される細菌又はファージプロモーターであるため, 当業者 であれば実際の実験要件に応じて構築物中のプロモーターを置き換えてタンパク質発現系に応用することにかかる技術的障害も存在しない。	
	向上	BLEU : 0.8224 (+0.1294) RIBES : 0.9775 (+0.0234)
第 20 回学習後 内容伝達 Lv : 5	上記 相違点 について,T7A2 などの各プロモーターはいずれも本分野で一般的に使用される細菌又はファージプロモーターであるため, 当業者 が実際の実験要件に応じて構築物中のプロモーターを置き換えてタン	

³⁶ エラーは 2 語だが、カウントは文単位につき 1 (文) である。一方、特許 NT では「**本領域技術人員**」は「**当業者**」と統一訳どおりに訳されたが、「**区別技術特徴**」が「**区別技術の特徴**」と誤訳された。特許誤訳に相当する箇所は 1 語だが、文単位のカウントとなるため、カウント上は汎用 NT と同じとなる。

	パク質発現系に応用することにはいかなる技術的障害も存在しない。	
同等	BLEU : 0.7397 (-0.0827) RIBES : 0.9692 (-0.0083)	

第10回、第15回、第20回学習後においても、審決用語「[区別技術特征](#)」「[本領域技術人員](#)」は安定して適訳語の「[相違点](#)」「[当業者](#)」が採用されている。なお、人手評価は学習5回ごとに実施したが、これら二語は人手評価対象外の回を含め第1回から第20回学習後までの全ての回で、両エンジンとも一切の訳ゆれなく完璧な安定を示した。本カテゴリのカウンタ結果と、これまで分析してきた各実例の状況から、評価用コーパスの大多数の文において同様の状況であると考えられる。つまり特許・審決用語に関しては、本事業の企図したとおり、中国審決コーパスデータの学習によって網羅的かつ安定的な翻訳精度向上が果たされたと結論できる。

ただし、カウンタ結果が示すとおり、学習を重ねても解消されなかった特許誤訳も一種のみ存在した。実例3-5(⇒3.4.5.)でも発生していた「[專利](#)」の訳し分け誤りである。その発生理由は同実例の分析時に既に考察したが、このエラーが、最終学習後まで残存した唯一の特許誤訳であった。

エラーのカテゴリ別カウンタでは、他の特許誤訳が学習の進展に伴い急速に改善されるなか、このエラーのみは解消せず、最終学習後も唯一の特許誤訳として残存した。具体的には、第5回学習後の時点で既に汎用NTの特許誤訳5件の全件、特許NTの特許誤訳も21件中17件が「[專利](#)」の訳し分け誤りであった。つまり、当時両エンジン間で特許誤訳のカウンタ数に大きな差が生じたのは、「[專利](#)」の訳し分け誤りが特許NTに偏って発生したためといえる。続く第10回学習後は同エラーは汎用NTで特許誤訳5件中の2件³⁷、特許NTで特許誤訳13件中の12件と引き続き特許NTに大きく偏って発生したが、大規模アップデート以降の第15回学習後は汎用NTが全15件に対し特許NTが全16件、第20回学習後は汎用NTの9件に対し特許NTが15件と、全ての特許誤訳が「[專利](#)」の訳し分け誤りのみになるとともに、両エンジンの発生件数にも大きな差は生じなくなっている。

ここで、「[專利](#)」の訳し分け誤りの発生状況を改めて実例で示しておく。サンプルには文番号#1365、特許NTの機械翻訳文を用いた。

【文番号：#1365、特許NT】

原文	证据 1-3、5-8 的公开时间均早于本专利的优先权日，可以作为现有技术评价本专利的创造性。
----	--

³⁷ 残り3件はそれまで適切に訳されていた用語・表現における突発的な誤訳や誤解釈、欠落であった。

正解訳文	証拠 1～3、5～8 の公開日がいずれも 本特許 の優先日より前であるため、従来技術として 本特許 の創造性を評価できる。	
無学習 内容伝達 Lv : 3	証拠 1-3、5-8 の開示時間はいずれも 本特許 の優先日より早く、従来技術で 本特許 の創造性を評価することができる。	
	—	BLEU : 0.4427 (—) RIBES : 0.8875 (—)
第 5 回学習後 内容伝達 Lv : 4	証拠 1-3、5-8 の公開日時はいずれも 本実用新案 の優先日より前であるため、 本実用新案 の創造性を評価するための従来技術として使用することができる。	
	向上	BLEU : 0.3185 (-0.1242) RIBES : 0.8379 (-0.0496)

本例は原文に「**本特許**」が二回出現する。この文では正解訳文のとおり「**本特許**」と訳すのが正しい。ただし、それはこの文が由来する中国審決の内容から判断できることであり、この文単独の記載内容からは「特許」と「実用新案」のどちらが正解かを判断することは不可能である。

こうした状況の中、無学習時の機械翻訳文では「**本特許**」という正しい訳語が採用されている。なお、両エンジンを通じて無学習時に「特許」が「実用新案」と訳された文は皆無であり、そもそもデフォルトの学習データ中には「特許」の訳語候補として「実用新案」は存在していなかったと見受けられる。このため無学習時には、「特許」を「実用新案」と訳すべき文は両エンジンとも全件が誤訳となっている。

これに対し、「特許」が「特許」と「実用新案」とに訳し分けられた中国審決コーパスデータを学習した結果、第 5 回学習後の翻訳文では、具体的な基準は不明ながら、文ごとに「特許」が「特許」と「実用新案」に訳し分けられるようになった。本例ではこれが裏目に出て「**本実用新案**」と誤訳になったが、コーパスの学習によって「特許」の訳語候補に「実用新案」という新たな選択肢が加わったことがわかる。

なお、もう一方のエンジンである汎用 NT で第 5 回学習後に採用された訳語は「**本特許**」であった。同じデータを学習させたにもかかわらず両エンジン間で訳語に違いが生じるという現象は、本報告書で定点観測した他の実例、他の用語でも数多く見られており、「みんなの自動翻訳」における訳語選定のランダム性を示すものといえる。

続いて、第 10 回学習後以降の特許 NT の機械翻訳結果を示す。

第 10 回学習後 内容伝達 Lv : 5	証拠 1-3、5-8 の公開日時はいずれも 本特許 の優先日より前であるため、従来技術として 本特許 の創造性を評価することができる。	
--------------------------	---	--

	向上	BLEU : 0.6217 (+0.3032) RIBES : 0.9208 (+0.0829)
第 15 回学習後 内容伝達 Lv : 5		証拠 1-3、5-8 の公開日時はいずれも 本特許 の優先日より前であるため、従来技術として 本特許 の創造性の評価に用いることができる。
	同等	BLEU : 0.6148 (-0.0069) RIBES : 0.9097 (-0.0111)
第 20 回学習後 内容伝達 Lv : 5		証拠 1-3、5-8 の公開日時はいずれも 本特許 の優先日より前であるため、従来技術として 本特許 の創造性を評価することができる。
	同等	BLEU : 0.6217 (+0.0069) RIBES : 0.9208 (+0.0111)

第 10 回学習後以降は、各回ともこの文における正解訳である「**本特許**」が選択された。汎用 NT も同じ状況である。ただし、前掲した全 100 文のカウント結果では、特許誤訳は第 5 回後以降も一定数発生しており、学習が進むにつれて「**専利**」の訳し分けの精度が改善したとは言い難い。本例の、人手評価対象外の回を含めた各回の履歴を詳細に追うと、大規模アップデート以前は特許 NT で第 1、3、5、11 回後、汎用 NT で第 8、11 回後と、両エンジンで散発的に「**本実用新案**」が採用されていたのに対し、アップデート以降は双方のエンジンとも全回「**本特許**」が採用されており、このことから、個々の文において採用される「**専利**」の訳語は大規模アップデート以降、一方の訳語に安定化した状況が見て取れる。ただし、評価対象文全体においては、大規模アップデート以降も「**専利**」の訳し分け誤りは引き続き一定量カウントされ続けていることから、文によっては誤った訳語で安定してしまった文も存在することがうかがえる。

「**専利**」の訳し分け誤りは本カウント結果が示すとおり最終学習後も残存しており、実用にあたっては何かしらの対処を講じることが望ましい。次章第 5.5.項においてさらに考察のうえ、対処策を提案する。

4.5.5.2. 「技術／一般誤訳」のカウント結果

技術用語・表現や一般用語・表現など、特許・審決用語以外の語句の誤訳は本カテゴリにカウントした。集計結果を下表に示す。

表 4.5.6. 各回学習前後の「技術／一般誤訳」該当文数カウント結果

エンジン	無学習	第 5 回学習	第 10 回学習	第 15 回学習	第 20 回学習
汎用 NT	40	21	18	11	13
	—	-19	-3	-7	+2
特許 NT	37	20	16	11	11
	—	-17	-4	-5	±0

本調査分析で学習に用いた中国審決コーパスデータは、デフォルト状態のエンジンではカバーされていない「審決書類特有の用語や表現、文型」に機械翻訳システムを適応させることを主目的としている。技術用語や一般用語も数多く含んでいるものの、全技術分野のあらゆる語彙を網羅しているわけではなく、したがって技術／一般用語の翻訳精度はエンジンの基礎学習に委ねられる度合いが大きい。よって本カテゴリは多分に「みんなの自動翻訳」それ自身の技術／一般用語の翻訳品質を測る意味合いを持つ。

上掲の集計結果を見ると、技術／一般誤訳は特許誤訳に比べて無学習の時点でのカウント数は半数以下と相対的に少ない。だが、それでも全体の 4 割程度の文において何かしらの誤訳が発生している。

また、無学習時から最終学習後まで全期間を通じ、汎用 NT と特許 NT のカウント数にほとんど差が生じていない。両エンジンは基礎学習データの一部が異なるが、この結果から、技術用語及び一般用語に関して、両エンジン間に実質的な品質差が存在しないことがうかがえる。

なお、技術／一般用語はエンジンの基礎学習データに大きく依存すると述べたが、実際のカウント結果では両エンジンとも学習を重ねるにつれてエラー数が漸減している。このように中国審決コーパスデータの学習によって解消された技術／一般誤訳の多くは、中国審決特有の文に含まれる一般用語・表現の改善であると考えられる。以下、具体例を示す。

【文番号：#769、特許 NT】

原文	因此，对于复审请求人的主张，合议组不予支持。	
正解訳文	よって，復審請求人の主張について，合議体は支持しない。	
無学習 内容伝達 Lv：1	従って、交渉要求者の主張に関しては、契約グループはサポートされない。	
	—	BLEU：0.0642 (－) RIBES：0.6866 (－)
第 5 回学習後 内容伝達 Lv：5	よって、復審請求人の主張については、合議体はこれを支持しない。	
	向上	BLEU：0.6515 (+0.5873) RIBES：0.9602 (+0.2736)
第 10 回学習後 内容伝達 Lv：5	よって、復審請求人の主張について、合議体は支持しない。	
	向上	BLEU：1.0000 (+0.3485) RIBES：1.0000 (+0.0398)
第 15 回学習後 内容伝達 Lv：5	よって、復審請求人の主張について、合議体はこれを支持しない。	
	同等	BLEU：0.7913 (-0.2087) RIBES：0.9726 (-0.0274)
第 20 回学習後 内容伝達 Lv：5	よって、復審請求人の主張について、合議体は支持しない。	
	同等	BLEU：1.0000 (+0.2087) RIBES：1.0000 (+0.0274)

本例は中国審決の頻出文であり、文そのものをひとつの審決表現と見なすこともできるが、カテゴリ別エラーカウントでは語句ごとにエラー判定しており、かつ文末の「不予支持」は審決以外でも使用される語であるため、一般用語・表現として扱われた。

「不予支持」は、この文では正解訳文のとおり「支持しない」と訳すのが適切である。無学習時の特許 NT の機械訳文ではこれを「サポートされない」と訳したため、誤訳と判定された。

前項の冒頭において、「(明細書に) サポートされていない」という言い回しは審決用語に準ずるため特許誤訳の対象とする旨を述べた。本例の「サポートされない／支持しない」も原語自体はこれと同じであるが、「主張を支持しない」という言い回しはごく一般的であり、「明細書にサポートされていない」のような審決特有の用法とはいえないため、特許誤訳ではなく一般誤訳に分類される。

なお、この誤訳は無学習時、すなわち中国審決コーパスデータ学習以前の機械訳文で発生しており、コーパス中の「明細書にサポートされていない」という言い回しが学習された結果ではない。したがって、少なくとも特許 NT の基礎学習データ中に「サポートする」という二つの訳語候補が存在していることがわかる。

この箇所は、中国審決コーパスデータを学習させた第 5 回学習後以降は各回とも「支持しない」と正しい訳語に改善され、誤訳は解消した。人手評価対象外の各回においても、

両エンジンとも第1回学習後から最終学習後まで全ての回でこの正しい訳語が採用されている。

中国審決コーパスデータでも「支持」(原語)は「支持する」と「サポートする」とに訳し分けられている。つまり、コーパス学習以降も「サポートする」という訳語は、有力候補の一つとして「支持する」とともに学習データ中に併存する状況である。このように同じ用語に対して複数の有力な訳語候補が存在する場合でも、本例のように特定の文脈で一方が多用されていれば、学習初期から適切な訳語が選択され、正しく訳し分けられることが示された。事実、自動評価時に事例3-1として定点観測した文番号#1(⇒3.4.1.)は「支持」(原語)を「サポートする」と訳すべき事例であったが、こちらも「請求項の保護範囲は明細書にサポートされている」という中国審決の定番表現に組み込まれていたため、コーパス学習後は常時「サポートする」と正しい訳し分けがなされ、「支持する」と誤訳されることはなかった。

このように、特許・審決表現に組み込まれた一般用語・表現は中国審決コーパスデータの学習により改善が進んだが、これに該当しない技術/一般用語は、中国審決コーパスデータ中に存在するか否かは偶然に委ねられており、含まれていない場合はコーパスを学習させても誤訳が改善されないことが多いと考えられる。次例はこれに該当する事例を示す。

【文番号：#122、汎用 NT】

原文	可见, 对比文件 4 给出了如何降低基础表达的技术手段, 且作用与其在本申请中为解决发明实际解决的技术问题所起的作用相同。	
正解訳文	このように、引用文献4には基礎発現をいかにして低減するかを技術的手段が示されており、その作用は本願の発明が実際に解決しようとする技術的課題を解決するための作用と同じである。	
無学習 内容伝達Lv:1		比較文書4はどのように基礎表現を低減する技術的手段を与え、その効果は本願で解決した技術的問題を解決するための役割と同じである。
	—	BLEU: 0.1778 (—) RIBES: 0.7791 (—)
第5回学習後 内容伝達Lv:3		以上から分かるように、引用文献4は、どのようにして基礎表現の技術的手段を削減するかを示しており、且つその作用が本願において発明が実際に解決しようとする技術的課題を解決するために果たす役割と同じである。
	向上	BLEU: 0.4713 (+0.2935) RIBES: 0.9003 (+0.1212)
第10回学習後	以上から分かるように、引用文献4は基礎表現をどのように低下させる	

内容伝達 Lv : 4	かという技術手段を与えており,且つその役割は本願において発明が実際に解決しようとする技術的課題を解決するために果たす役割と同じである。
	向上 BLEU : 0.4609 (-0.0104) RIBES : 0.8997 (-0.0006)
第 15 回学習後 内容伝達 Lv : 4	以上から分かるように,引用文献 4 ほどのように基礎発現を低下させるかという技術手段を与えており,且つ作用はそれが本願において発明が実際に解決しようとする技術的課題を解決するために果たす作用と同じである。
	同等 BLEU : 0.4857 (+0.0248) RIBES : 0.9074 (+0.0077)
第 20 回学習後 内容伝達 Lv : 4	このことから分かるように,引用文献 4 ほどのようにして基礎表現を減少させるかという技術手段を与えており,且つ作用は本願において発明が実際に解決しようとする技術的課題を解決するために果たす作用と同じである。
	向上 BLEU : 0.4673 (+0.2699) RIBES : 0.8801 (+0.0846)

本例では、「基礎表达／基礎発現」という技術用語に注目する。各回の訳語の推移を見ると、汎用 NT では無学習時、第 5 回学習後、第 10 回学習後と各回「基礎表現」と訳され、これが技術誤訳と判定されていたが、第 15 回学習後には「基礎発現」と適訳となり、誤訳が解消された。しかし第 20 回学習後は再び「基礎表現」が採用されている。

この推移のみを見ると、第 11 回から第 15 回にかけて追加した学習データ中に「基礎表达」を「基礎発現」と訳している文例が含まれており、第 15 回後の「基礎発現」への改善はこれが反映された結果のように映る。だが実際には、この期間の追加学習データには「基礎表达」という語は含まれていない。また、両エンジンの人手評価対象外を含めた各回の機械翻訳結果を詳細に調査したところ、下表に示すとおり、「基礎発現」は無学習時から訳語候補のひとつとして散発的に採用されていることが判明した。つまり、上掲の第 15 回後の「基礎発現」への改善は、「みんなの自動翻訳」における訳語選定のランダム的な選択の結果にすぎない。

表 4.5.7. 文番号#122 における各回学習前後の「基礎表达」の訳語

	汎用 NT	特許 NT
#0	基礎表現	基礎発現
#1	基礎表現	基本的な発現
#2	基礎表現	基礎的表現
#3	基礎表現	基礎表現
#4	基礎発現	基本的な発現

#5	基礎表現	基本発現
#6	基礎発現	基本的に発現する
#7	ベース表現	ベース発現
#8	基礎発現	基礎表現
#9	基礎表現	ベース表現
#10	基礎表現	基礎表現
#11	ベース表現	基礎表現
#12	基礎表現	基礎的な表現
#13	基礎表現	基礎発現
#14	基礎表現	基礎表現
#15	基礎発現	基礎的表現
#16	基礎表現	表現の基礎
#17	基礎的な表現	基礎表現
#18	基礎発現	基礎表現
#19	基礎表現	基礎表現
#20	基礎表現	基礎表現

こうした訳語のランダム的な変化は、第12回学習前の大規模アップデートを境界として前後の訳語を比べると、アップデート以降はある程度抑制されていることが分かる。

「ベース表現」のようなマイナーな不適訳語が不用意に採用されることが少なくなり、両エンジンとも多くの回で「基礎表現」という訳語が採用されるようになった。この訳語が採用されない回も、用語前半の「基礎」部分で「基礎的」、後半の「表达」部分で「発現」という訳語が採用されるのみであり、例えばアップデート以前に散発的に採用されていた「基礎：ベース」や「基礎：基本(的)」という訳語はアップデート以降は採用されなくなった。

こうした推移を総括すると、以下のような状況と推察される：本例の「基礎表达」に関しては、学習データ中に訳例が存在せず、このため「基礎」と「表达」とに分解して翻訳されている。学習データ中の訳語候補としては少なくとも前者に対して「基礎」「基礎的」「ベース」「基本」「基本的」の5種、後者に対して「発現」と「表現」の2種が存在しており、大規模アップデート以前は学習回ごとにこれらの候補からランダム的に訳語が採用されていたが、大規模アップデート以降は、訳語選定ロジックの改良により、学習データ中で有力な訳語候補が安定して採用されるようになり、その結果、「基礎」に関しては適訳「基礎」が最有力候補として安定的に採用されるようになり、「ベース」「基本」「基本的」といった非有力な訳語候補は採用されなくなった。「表达」においても訳語は安定したが、こちらは本例における適訳の「発現」ではなく「表現」のほうが最有力候補となっ

たため、「基礎表达」の単位では「基礎表現」という不適訳語が多くの回で採用される結果となった。ただし、アップデート以降も「基礎：基礎的」や「表达：発現」という次点の訳語候補が突発的に採用される回も存在し、こうした訳語の不安定性が「みんなの自動翻訳」の採用するニューラル翻訳方式そのものに内包される性質であることも同時に示された。

もともと一般／技術用語は、特許・審決用語と異なり、特に単語レベルでは複数の訳語に訳し分けられることが多い。本例に関しても、「基礎表达」の単位であれば適訳語は「基礎発現」となるが、「基礎」「表达」という個々の単語レベルでは、前者ならば「基本」、後者ならば「表現」という訳語にも多くの使用例がある。学習データ中に「基礎表达」の使用例がないため、「基礎」と「表达」の訳語にそれぞれ「基礎」「発現」を安定して選定するだけの根拠が得られないものと考えられる。

したがって、本例の技術誤訳は、基礎学習データを含めた学習データ中に「基礎表达」という複合語を用いた十分な文例が存在しないことが原因と推察される。つまり、根本的には学習データにおける技術用語の網羅性の問題といえる。

ニューラル翻訳の場合、「基礎表达」での訳例を学習データに加えれば必ずその訳語が採用されるという保証はないが、コーパス学習後も適訳語が採用されない用語については、ピンポイントで学習データを作成・補充することが最も改善の可能性が高い方策であると考えられる³⁸。

³⁸ 本調査分析で最終学習後も技術／一般誤訳が検出された文は、評価用コーパスに用いたため学習データからは除外していたが、実際には中国審決コーパスデータの一部である。このため、これらの文で技術／一般誤訳となっていた語も、コーパス全件を学習させれば適訳語に改善される可能性がある。

4.5.5.3. 「未知語」のカウント結果

原文の中国語（又はその一部）が、和訳されずそのまま中国語で出力されている場合、「未知語」と見なした。

なお中日翻訳では、語単位で翻訳することができなかった場合、その語が一文字単位に分解されて中国語漢字から日本語漢字へ変換されることがある。その結果、一語中に日本語文字と中国語漢字が混在する場合があります（例：「从属权利要求」⇒「从属請求項」。正しくは「従属請求項」）。また、用語の各文字が日本語漢字に置き換わっているだけで、日本語として意味をなさないような場合も想定される。本評価では、これらのケースも全て未知語エラーと判定した。カウント結果は以下のとおりである。

表 4.5.8. 各回学習前後の「未知語」該当文数カウント結果

エンジン	無学習	第 5 回学習	第 10 回学習	第 15 回学習	第 20 回学習
汎用 NT	19	1	1	1	0
	—	-18	±0	±0	-1
特許 NT	2	1	1	1	0
	—	-1	±0	±0	-1

上表のとおり、未知語の発生は無学習時の汎用 NT で 19 文とやや多かったが、検出された未知語はほぼ「合議体」を意味する「**合議組**」と「**技術的解決手段**」を意味する「**技術方案**³⁹」の 2 語に絞られていた。両語とも中国審決では頻出の用語であるためカウント数は伸びたが、特許・審決用語が学習初期から正解訳文どおりの適訳語で安定するという傾向どおり、初回学習後にはこれら 2 語に関する未知語エラーは全て解消している。それ以降は、文番号#1146 の「**舒焊**」のみが唯一の未知語として両エンジンで第 15 回時まで毎回カウントされていたが、この語に関しても第 20 回学習後には両エンジンで訳出されるようになり、結果、未知語のカウントはゼロとなった。以下、この語を含む文番号#1146 の、汎用 NT による第 15 回後と第 20 回後の機械翻訳文を示す。

³⁹ 原語は「技术方案」であり、翻訳文では「技術方案」と日本の漢字に変化されていたが、単に中国漢字を対応する日本漢字に置換しただけで日本語として意味をなさないため、事実上中国語のまま見なし、未知語と判定した。日英翻訳において日本語がローマ字で出力されるケースと同じ扱いである。ただし中日翻訳の場合、漢字が表意文字であることから「日本語として意味をなすか否か」は微妙な判定となることも多く、誤訳か未知語かの線引きは難しかった。

【文番号：#1146、汎用 NT】

原文	<p>(2) 对比文件 3 与权利要求 1 的安装结构不同, 对比文件 3 电机壳体通过膏状舒焊涂敷能够使平坦状底面部粘着在印刷配线板的固定用图形上进行固定, 是通过电机壳体直接将电机固定在印刷配线板上, 没有相应的独立于电机壳体之外的安装件, 而本专利权利要求 1 中是通过电机主体以外的安装件固定电机, 因此对比文件 3 没有公开安装件以及定位凹部、插入支持片和底部提高平板状部能够收纳在所述马达壳体的粗度以内”。</p>	
正解訳文	<p>(2) 引用文献 3 と請求項 1 の取付け構造は異なり, 引用文献 3 のモータケースはペースト状半田コーティングによって平坦状底面部をプリント基板の固定用パターン上に接着固定して, モータケースによって直接モータをプリント基板上に固定するものであり, 相応のモータケースの外の独立した据付金具はなく, 更に本特許請求項 1 はモータ本体以外の据付金具を介してモータを固定するものであり, よって引用文献 3 は据付金具及び位置決め凹部、差し込み支持片及び底上げ平板状部が前記モータケースの太さ以内に収納できることを開示していない。</p>	
第 15 回学習後 内容伝達 Lv : 3	<p>(2)引用文献 3 は請求項 1 の取り付け構造とは異なり,引用文献 3 のモータケースはペースト状の舒焊部を塗布することによって平坦状底面部をプリント配線板の固定用パターンに接着して固定することができ,それはモータケースを介してモータをプリント配線板に直接固定するものであり,対応するモータケースから独立した取り付け部材がなく,本実用新案の請求項 1 ではモータ本体以外の取り付け部材を介してモータを固定するため,引用文献 3 には取り付け部材及び位置決め凹部、挿入支持片及び底上げ平板部が前記モータケースの太さ以内に収納できることが開示されていない」。</p>	
	—	BLEU : 0.3536 (–) RIBES : 0.8132 (–)
第 20 回学習後 内容伝達 Lv : 4	<p>(2)引用文献 3 は請求項 1 の取り付け構造とは異なり,引用文献 3 のモータハウジングはペースト状のはんだ付け塗布により平坦状底面部をプリント配線板の固定用パターンに接着して固定することができ,モータハウジングを介してモータをプリント配線板に直接固定するものであり,モータハウジングから独立した対応する取付具がないが,本特許の請求項 1 ではモータ本体以外の取付具によりモータを固定するため,引用文献 3 には取付具及び位置決め凹部、挿入支持片及び底上げ平板状部が前記モータハウジングの粗さ以内に収納できることが開示されていない」。</p>	
	向上	BLEU : 0.4163 (+0.0627) RIBES : 0.8216 (+0.0084)

懸案の「**舒焊**」は正解訳文にあるとおり「**半田コーティング**」を意味する語だが、中国語特有の省略がなされており、そのせいで前回時までは汎用 NT の機械訳文に見られるとおり（「**舒焊部**」）、両システムで常時、中国語がそのまま出力されていた。これに対し、今回の汎用 NT では「**はんだ付け塗布**」と、不自然ではあるが十分に意味が推察できる日本語で訳出されるようになり、未知語が解消された。なお、特許 NT においても第 20 回学習後の訳文では「**舒焊**」は「**はんだ塗布**」と訳出されており、汎用 NT よりも良好な訳語となっている。

人手評価対象外の各学習回の訳文を含めて詳しく調査したところ、本例では「**舒焊**」という語は第 17 回学習後以降は、例えば「**溶接塗布**」（第 17 回・特許）、「**シーム溶接**」（第 18 回・汎用）など両エンジンで日本語に翻訳されるようになり、第 19 回後からは「**はんだ／半田**」という本来の意味で訳されるようになっていた。この時期に学習させた中国審決コーパスデータにも「**舒焊**」を含む文は存在していないが、何らかの学習データにより「**舒**」と「**焊**」をそれぞれ単独で日本語に解釈することを学習したものと推察される。

このように、人手評価対象 100 文においては、「**舒焊**」が訳出されるようになったことで、最終学習後は未知語エラーは一切顕出されなくなった。無学習の時点から汎用 NT の「**合议组**」と「**技術方案**」の二種にほぼ限定されていたことを考えると、「みんなの自動翻訳」自体、未知語エラーがあまり発生しない性質のエンジンであるといえる。ただし、これは必ずしも「原文の全ての語が（正否は別として）確実に訳出される」ことを意味しない。例えば第 3.4.5.項の実例 3-5 で見られたように、「みんなの自動翻訳」では、翻訳不能な用語・表現は、未知語として出力されるのではなく、訳文から欠落してしまう場合もある。この場合、エラーカウント上は別種の「欠落エラー」としてカウントされる。欠落エラーの発生状況は、第 4.5.5.6.項において後述する。

4.5.5.4. 「係り受けエラー」のカウント結果

文中の要素間の関係性（係り受け）が原文と異なっている場合は、係り受けエラーと見なした。カウント結果は以下のとおりである。

表 4.5.9. 各回学習前後の「係り受けエラー」該当文数カウント結果

エンジン	無学習	第 5 回学習	第 10 回学習	第 15 回学習	第 20 回学習
汎用 NT	42	19	17	7	5
	—	-23	-2	-10	-2
特許 NT	49	15	20	4	6
	—	-34	+5	-16	+2

ニューラル機械翻訳は文法規則を考慮して翻訳文を構築する方式ではなく、文中の係り受けも、学習データの内容を解析して、最も確からしい語順で翻訳文を組み立てた結果である。中国審決コーパスデータはあらゆる文構造やフレーズを網羅しているわけではないため、各文の係り受けにもコーパスから学習できるものとできないものが混在する。つまりはデフォルトの学習内容に委ねられる部分が大きく、その意味において技術／一般誤訳に近い条件のエラーといえる。

無学習時は両エンジンとも半数弱の文で何らかの係り受けエラーが発生していたが、中国審決コーパスの第 5 回学習後には汎用 NT で 19 文、特許 NT で 15 文と大幅に改善された。これらの多くは「コーパスから学習できた文構造・フレーズ」に属するものであり、中国審決に特有かつ頻出の文が中心となる。以下、一例を示す。

【文番号：#468、汎用 NT】

原文	撤销国家知识产权局于 2016 年 10 月 09 日对本申请作出的驳回决定。	
正解訳文	国家知識産権局が 2016 年 10 月 9 日に本願に対して行った拒絶査定を取り消す。	
無学習 内容伝達 Lv : 1	国家知的財産権局の廃止は 2016 年 10 月 09 日に本申請に対する棄却決定を下した。	
	—	BLEU : 0.1009 (—) RIBES : 0.8165 (—)
第 5 回学習後 内容伝達 Lv : 5	国家知識産権局が 2016 年 10 月 09 日に本願に対し行った拒絶査定を取り消す。	
	向上	BLEU : 0.7242 (+0.6233) RIBES : 0.9753 (+0.1588)
第 10 回学習後 内容伝達 Lv : 5	国家知識産権局が 2016 年 10 月 09 日に本願に対し行った拒絶査定を取り消す。	
	同等	BLEU : 0.7242 (±0) RIBES : 0.9753 (±0)

第 15 回学習後 内容伝達 Lv : 5	国家知識産権局が 2016 年 10 月 09 日に本願に対し行った拒絶査定を 取り消す。	
	同等	BLEU : 0.7242 (±0) RIBES : 0.9753 (±0)
第 20 回学習後 内容伝達 Lv : 4	国家知識産権局が 2016 年 10 月 09 日に本願に対し行った拒絶査定を 取り消す。	
	同等	BLEU : 0.7242 (±0) RIBES : 0.9753 (±0)

無学習時の翻訳文は原文の「撤銷」の係り先や品詞解釈が誤っており、ナンセンスな内容になっている。中国語文法的にも無理のある解釈である。

原文は単純な構造の短文であるが、中国審決特有の文であり、デフォルト学習データに類似文が存在していなかったと考えられる。ニューラル機械翻訳は文法規則を用いて翻訳を行う方式ではないため、文法的には単純な構造の文であっても、類似の文例が学習データ中にある程度存在しないと、本例のように不正確な係り受けの翻訳文となってしまうこともあるようである。中国審決で使用されている文法ルール自体は一般の文と共通であるが、それにもかかわらず無学習時に半数近い文で係り受けエラーが多発していた理由は、こうしたニューラル翻訳方式の方法論に起因していると考えられる。

これに対し、中国審決コーパスデータ学習後は、上掲のとおり本例は全ての回で正しい係り受けで翻訳されるようになった。人手評価対象外の全ての回、そして特許 NT まで範囲を広げても同じ状況である⁴⁰。この結果は、「みんなの自動翻訳」を始めとするニューラル機械翻訳における係り受けの精度は、文法的な構造の単純さよりも、原文を構成する各語句が学習データ中に豊富に存在するか否かに大きく左右されるという、前述の推論と整合している。そして上掲のカウント結果で第 5 回学習後に係り受けエラーの数が大幅に減少したのも、中国審決コーパスデータを介して審決用語・表現を含む文例が大量に学習された結果と見なせる。

その後、第 10 回学習後の係り受けエラーの発生数は第 5 回後とおおむね同程度で、特段の改善は見られなかった。上記推論からは、審決用語・表現がひとつおりの学習された後も係り受けエラーが引き続き発生している文の多くは、学習データ中に参考となる文例が存在しない文と考えられる。ただし、それに加えて、学習データ中に参考となる文例が存在していたとしても、その文例からどのような情報を取得して翻訳文を構築するかはエンジンのロジック次第であり、文によってはその過程で係り受けを誤る可能性も考えられる。実際、この時期の「みんなの自動翻訳」では、学習データ追加により翻訳文が不安定

⁴⁰ 「本願に対し」が「本願に対して」と変化した回が数回あるのみである。

に変化する傾向が強く、その結果、前回学習時には正しかった係り受けが突発的に悪化するケースも散見されており、一部こうした文もカウントされていると考えられる。

このように係り受けエラーの原因ともなり得る翻訳文の不安定な変化は、第12回学習前の大規模アップデート以降、かなり抑制された。アップデート時にエンジンのロジックが見直され、より学習データに準拠した訳語選定がなされるようになった結果であるが、その効果は翻訳文の語順、すなわち係り受けの精度にも及んでいる。第10回学習後には停滞していた係り受けエラーのカウント数が第15回学習後になって再び顕著に減少したのも、この大規模アップデート時のロジックの改良によるものと考えられる。その後、第20回学習後にカウントされた係り受けエラーも第15回後とほぼ同数にとどまったことから、当時の改善が突発的な誤訳の一時的な減少によるものではなく、大規模アップデート以降の平均的な発生数であることも確認された。

ニューラル機械翻訳では、翻訳文の語順は文法ルールではなく、入力文を構成する各語句について学習データから取得した情報に基づいて決定される。係り受けエラーはこの語順決定の結果として生じるものであり、したがって係り受けエラーの改善には学習データにおける語句の網羅性が重要となる。本カウント結果からは、審決関連語の文例の不足による係り受けエラーは第5回学習まででおおむね解消されたと考えられる。その後、エンジンの改良により翻訳文構築の精度も改善され、係り受けエラーの発生はごく少量に抑えられるようになった。引き続き残留するエラーは、技術／一般誤訳と同様、文例の不足が原因であり、ピンポイントで対処することで改善される可能性がある。

4.5.5.5. 「日本語表現エラー」のカウント結果

本カテゴリには、誤訳ではないが日本語として不自然、読みにくい翻訳となっているものをカウントした。例えば不適な接続詞や舌足らずな表現などによって文意が曖昧、説明不足、ミスリーディングになっているようなケースなどが該当する。以下、集計結果を示す。

表 4.5.10. 各回学習前後の「日本語表現エラー」該当文数カウント結果

エンジン	無学習	第 5 回学習	第 10 回学習	第 15 回学習	第 20 回学習
汎用 NT	43	24	23	13	10
	—	-19	-1	-10	-3
特許 NT	42	18	21	17	14
	—	-24	+3	-4	-3

本エラーも技術／一般誤訳や係り受けエラーと同様、無学習時は半数弱の文で検出され、第 5 回学習後には顕著に改善したが、続く第 10 回後にはほとんど改善が見られなくなっている。第 5 回までに審決特有の文や言い回しにおける不自然な日本語表現が改善され、その一方で、審決表現との関連が薄くコーパス中に参考となる文例が乏しいものは改善されずに残存した結果と考えられる。

以下、審決特有の文における日本語表現エラーの改善事例を二例示す。二例とも、第 10 回学習後以降の機械翻訳結果は第 5 回後と実質的に同じであるため掲載は省略する。

【文番号：#784、汎用 NT】

原文	申請人（下称复审请求人）对上述驳回决定不服，于 2016 年 9 月 14 日向专利复审委员会提出了复审请求，并未修改申请文件。				
正解訳文	出願人（以下、復審請求人という）は上記拒絶査定を不服とし、2016 年 9 月 14 日に特許復審委員会へ復審請求を行ったが、出願書類に対する修正はしていない。				
無学習 内容伝達 Lv：3	申請者(以下,復審請求者)は,上記棄却決定に対して不服となり,2016 年 9 月 14 日に復審請求を特許復審委員会に提出し,申請文書を修正しなかった。				
	—	BLEU：0.3034（－） RIBES：0.7344（－）			
第 5 回学習後 内容伝達 Lv：5	出願人(以下、復審請求人という)は上記拒絶査定を不服とし,2016 年 9 月 14 日に特許復審委員会へ復審請求を行ったが,出願書類に対する修正はしていない。				
	向上	BLEU：1.0000（+0.6966） RIBES：1.0000（+0.2656）			

本例の無学習時の機械翻訳文は、冒頭部が「申請者は棄却決定に対して不服となり」となっている。係り受けが誤っているというよりは、日本語表現が稚拙、不自然であると見なされ、日本語表現エラーにカウントされた。「不服」(原語)自体は一般的な語であるが、おそらくは目的語である「驳回決定」がデフォルト学習データ中に存在せず、この語が目的語となった際に「不服」をどう訳すべきかという情報が不足していたものと推察される。中国審決では最頻出の表現であり、コーパスの学習により豊富な訳例が得られた結果、第5回学習後には「出願人は上記拒絶査定を不服とし」と正解訳文どおりの自然な日本語表現に改善された。

【文番号：#1011、汎用 NT】

原文	在此基础上，合议组充分听取了双方意见。	
正解訳文	これを基礎として，合議体は双方の意見を十分に聴取した。	
無学習	それに基づいて，议组は双方の意見を十分に聞きました。	
内容伝達Lv：3	－	BLEU：0.3634 (－) RIBES：0.8480 (－)
第5回学習後	これに基づき，合議体は双方の意見を十分に聴取した。	
内容伝達Lv：5	向上	BLEU：0.7725 (+0.4091) RIBES：0.9635 (+0.1155)

この事例では無学習時、「聞きました」と「ですます調」が採用された。人手評価対象文全100文、二種の翻訳エンジン、各回の評価を通じ、ですます調が採用されたのはこの一回のみである。意味や係り受けは間違っておらず、唐突な「ですます調」が周囲から浮いていて不自然であるにすぎないため、日本語表現エラーと判定された。こちらも初回学習の時点で早々に解消され、以後再発はしていない。

上記二例はどちらも中国審決における頻出の文パターンに属しており、文全体が中国審決の頻出表現とも見なせる。よって本質的にはむしろ審決用語の誤訳に近く、このため学習初期から正解訳文どおりの適訳語で安定するという審決用語の傾向どおりの結果となった。つまり、本事業で改善を図った中国審決表現における日本語表現エラーは、これら二例に代表されるように、中国審決コーパスデータの学習によりおおむね改善されたと考えられる。

日本語表現エラーは当然ながら審決特有表現と無関係な箇所でも発生するが、これらは係り受けエラーと同じく、学習データ中に十分な文例が存在しない語句が文中に含まれているか、若しくは学習データ中に文例は存在するものの、何らかの理由でエンジンが翻訳文を正しく構築できなかったことが原因と考えられる。このため、中国審決コーパスデータをさらに増強したとしても、それが改善に直結する可能性は低い。むしろ、改善が必要な日本語表現エラーをピンポイントで学習させるほうが効率的に改善を果たせる可能性が高いであろう。

4.5.5.6. 「欠落エラー」のカウント結果

原文中に存在する語や表現が翻訳文から消失してしまっている場合は「欠落エラー」にカウントした。ただし、冗長な原文における意図的な省略など、翻訳上適切かつ原文の内容を損なわない措置と見なされるものは除いた。以下、欠落エラーの各回のカウント結果を示す。

表 4.5.11. 各回学習前後の「欠落」該当文数カウント結果

エンジン	無学習	第 5 回学習	第 10 回学習	第 15 回学習	第 20 回学習
汎用 NT	38	10	15	2	4
	—	-28	+5	-13	+2
特許 NT	45	13	13	1	2
	—	-32	±0	-12	+1

欠落エラーは中国審決コーパスデータの学習を経た第 5 回学習後に大きく減少し、第 10 回後には停滞したが、「みんなの自動翻訳」の大規模アップデートを経た第 15 回学習後に再び大きく減少し、以降はほとんど発生しなくなった。

エラーが検出された各文の状況を見ると、学習データ中にその用語が存在していないために欠落したと見なされるタイプと、過去学習回では訳出できており、したがって学習データ中に確実に存在する用語がエンジンの構文解析結果の変化などによって突発的に欠落したと見なせるタイプとの大別できる。

無学習時、すなわち中国審決コーパスデータ学習前の機械翻訳結果では、審決用語関連の欠落エラーが多発していた。これらが「学習データ中に用語が存在していないために欠落したと見なされるタイプ」に相当する。以下、一例を示す。

【文番号：#881、汎用 NT】

原文	并于 2014 年 7 月 14 日，向双方当事人发出无效宣告请求口头审理通知书，告知本案定于 2014 年 8 月 22 日进行口头审理。	
正解訳文	かつ 2014 年 7 月 14 日に，双方の当事者へ無効宣告請求の口頭審理通知書を送付し，本案に対し 2014 年 8 月 22 日に口頭審理を行うことを通知した。	
無学習 内容伝達 Lv : 2	また、2014 年 7 月 14 日に双方の当事者に対して口頭審理通知書を発行し、2014 年 8 月 22 日に口頭審理を行うことを提案した。	
	—	BLEU : 0.5135 (-) RIBES : 0.9317 (-)

第 5 回学習後 内容伝達 Lv : 4	且つ 2014 年 7 月 14 日に,双方の当事者へ無効宣告請求の口頭審理通知書を送付し,本案は 2014 年 8 月 22 日に口頭審理を行うことを告知した。	
	向上	BLEU : 0.8481 (+0.3346) RIBES : 0.9816 (+0.0499)
第 10 回学習後 内容伝達 Lv : 5	且つ 2014 年 7 月 14 日に,双方の当事者へ無効宣告請求の口頭審理通知書を送付し,本案に対して 2014 年 8 月 22 日に口頭審理を行うことを通知した。	
	向上	BLEU : 0.9123 (+0.0642) RIBES : 0.9879 (+0.0063)
第 15 回学習後 内容伝達 Lv : 4	且つ 2014 年 7 月 14 日に,双方の当事者へ無効宣告請求口頭審理通知書を送付し,本案が 2014 年 8 月 22 日に口頭審理を行うことを定めたことを通知した。	
	劣化	BLEU : 0.7829 (-0.1294) RIBES : 0.9734 (-0.0145)
第 20 回学習後 内容伝達 Lv : 5	且つ 2014 年 7 月 14 日に,双方の当事者へ無効宣告請求の口頭審理通知書を送付し,本案について 2014 年 8 月 22 日に口頭審理を行うことを通知した。	
	向上	BLEU : 0.9123 (+0.1294) RIBES : 0.9879 (+0.0145)

本例では、審決用語である「无效宣告请求（無効宣告請求）」が無学習時の翻訳文から完全に消失している⁴¹。中国審決では頻出の語だが、一般の文書で使われることはまれであり、デフォルト学習データ中には文例が存在していない可能性が高い。

無学習時の翻訳文では他にも文番号#123の「创造性劳动（創造的な労力）」、#407の「区别在于：（相違点は以下のとおりである：）」、#785、#1806、#1890の「复审请求人（復審請求人）」、#1080の「实施细则（実施細則）」、#1520、#1521の「专利权人（特許権者）」、#1651、#1657、#1745の「合议组（合議体）」、#1893の「审查指南（審査基準）」といった審決用語の欠落が検出されている。これら以外の欠落エラーも、例えば「认为：（主張は以下のとおりである：）」など、一般用語が審決特有の用いられ方をされているものが多い。

このタイプの欠落エラーは審決用語・表現が学習されていないことが原因であるため、中国審決コーパスデータを学習させることで改善することが多い。本例も、学習を経た第5回後以降は各回とも正解訳文どおり「無効宣告請求」と訳出されるようになり、欠落エラーは解消している。人手評価対象外の各学習回まで範囲を広げても、コーパス学習後は

⁴¹ なお本例の無学習時の翻訳文では原文の「本案」に対応する訳語も訳出されていないが、こちらは欠落か意図的な省略かは微妙である（評価者は欠落と判定）。

全 20 回×2 エンジンの全回で正しく「無効宣告請求」と訳出されており、欠落の原因が、審決用語が未学習であったためとわかる。上掲のカウンtr結果では、第 5 回学習後に欠落エラーのカウンtrが大幅に減少しており、こうした審決用語が未学習であることが原因で発生していた欠落エラーが、コーパスの学習によって一挙に解消されたことを示している。

前述の未知語エラーの考察 (⇒4.5.5.3.) においては、「みんなの自動翻訳」で未知語がさほど発生していない理由として、未学習の語の多くは未知語として出力されるのではなく翻訳文から欠落してしまっているという可能性を示唆した。本例を始め多くの審決用語が無学習時に欠落していたことから、この推測はひとまず肯定される。

ただし、未学習の用語であっても必ず翻訳文から欠落するわけではないということも、本例の分析により判明した。本例の「無効宣告請求」は、特許 NT では無学習時も「無効な告知要求」と単語に分解して訳出されており、誤訳ではあるが欠落はしていない。同じ文の同じ用語であるのにエンジン間で扱いが不安定に変化するの「みんなの自動翻訳」の傾向であるが、未学習の用語は、あくまで欠落しやすい傾向であるにすぎない。

「みんなの自動翻訳」で見られる欠落エラーにはもう一種、「過去学習回では訳出できていた用語が突発的に欠落するタイプ」も存在する。例えば第 3.4.2.項の実例 3-3 における「破乳步骤 (解乳化ステップ)」の第 1 回学習後と第 3 回学習後における欠落や、第 3.4.4.項の実例 3-4 における「匹配的 (見合う)」の第 13 回学習後のみでの突発的な欠落もこちらのタイプに該当する。このタイプの典型的な実例を以下に示す。

【文番号：#1010、汎用 NT】

原文	专利权人认为：权利要求 1-3 相对于证据 2 和公知常识的结合不具备创造性属于新增加的理由，且已经超出了规定的期限，应当不予考虑。	
正解訳文	特許権者の主張は以下のとおりである：請求項 1～3 が証拠 2 及び技術常識の組み合わせと比較して創造性を有しないとの主張は新たに追加された理由に該当し、且つ所定の期限を過ぎているため、考慮すべきではない。	
無学習 内容伝達 Lv：1	特許権者は、1-3 の権利が証拠 2 と公常識の結合に対して創造性が新たに増加した理由ではなく、規定の期限を超えたことを考慮して、考慮すべきであると考えている。	
	—	BLEU：0.1121 (—) RIBES：0.7182 (—)

第 5 回学習後 内容伝達 Lv : 5	特許権者の主張は以下のとおりである:請求項 1-3 は証拠 2 と技術常識の組み合わせに対して創造性を有さないことは新たに追加された理由に属し,且つ規定の期限を超えているため,考慮すべきではない。	
	向上	BLEU : 0.5648 (+0.4527) RIBES : 0.9444 (+0.2262)
第 10 回学習後 内容伝達 Lv : 2	実用新案権者の主張は以下のとおりである:請求項 1-3 は証拠 2 と技術常識の組み合わせに対して創造性を有さず,且つ規定された期限を過ぎており,考慮すべきではない。	
	劣化	BLEU : 0.4100 (-0.1548) RIBES : 0.9109 (-0.0335)
第 15 回学習後 内容伝達 Lv : 4	実用新案権者の主張は以下のとおりである:請求項 1-3 が証拠 2 と技術常識の組み合わせに対して創造性を有さないことは新たに追加された理由に属し,且つ規定期限を過ぎていているため,考慮すべきではない。	
	向上	BLEU : 0.5998 (+0.1898) RIBES : 0.9444 (+0.0335)
第 20 回学習後 内容伝達 Lv : 4	実用新案権者の主張は以下のとおりである:請求項 1-3 が証拠 2 と技術常識の組み合わせに対して創造性を有さないことは新たに追加された理由であり,且つ規定の期限を過ぎていているため,考慮すべきではない。	
	同等	BLEU : 0.6006 (+0.0008) RIBES : 0.9422 (-0.0022)

本例における「属于新増加的理由 (新たに追加された理由に該当し)」は、一般用語ではあるが審決特有の言い回しで、審決表現に準ずる。このため中国審決コーパスデータを学習させた第 5 回学習後には「新たに追加された理由に属し」と、他の多くの審決用語と同様、正解訳文と同義の適切な訳語に改善されている⁴²。

これに対し、第 10 回学習後の翻訳文ではこの文言が完全に欠落してしまっている。過去回で訳出されている以上、無学習時の審決用語のように学習データ中に存在しないということはありえず、明らかに「過去学習回では訳出できていた用語が突発的に欠落するタイプ」、すなわち学習データ中に存在する用語が、エンジンの構文解析結果の変化などによって突発的に欠落してしまうタイプの欠落エラーに属している。実際、両エンジン、全学習回を通じ、本例でこの箇所の欠落が生じたのはこの 1 回のみであった。

こちらのタイプの欠落エラーは、学習データの不足ではなく、エンジンの構文解析が何らかの理由で突発的に失敗してしまうことが原因と推察される。このため同じ文で継続的に発生することは少ないが、エンジンのロジックに起因するため学習データの追加による改

⁴² 無学習時の翻訳文も欠落エラーは生じていないが、「新たに増加した理由ではなく」と正反対の意味に訳されてしまっている。これは文末の「考慮すべきでない」に係るべき否定を誤った係り受けエラーである。

善は難しく、毎回同程度に発生していると考えられる。前掲のカウンtr結果において第5回後から第10回後にかけてエラー数が減少しなかったのも、このタイプの欠落エラーが常時一定規模発生しているためである。

だが、続く第15回学習後にはエラーのカウンtr数が顕著に減少し、汎用NTでは2文、特許NTでは1文とほとんど発生しなくなった。第20回学習後も、汎用NTが4件、特許NTでは2件と誤差レベルの微増にとどまっている。これらの結果は、上記「過去学習回では訳出できていた用語が突発的に欠落するタイプ」の欠落エラーについても、大規模アップデート以降はほぼ解消されたことを意味している。学習データの追加で改善する性質のエラーではないため、これは第12回時の「みんなの自動翻訳」の大規模アップデートにおいて、エンジンのロジックに改良が加えられたものと推察される。

最終学習後に検出された欠落エラーは、第4.5.4.4項で取り上げた「内容伝達レベルが最も低かった実例」で低評価の原点となった「软硬两项 (#1521)」や「企业范围的 (#1890)」など、翻訳難易度が高い技術／一般用語における「学習データ中に用語が存在していないために欠落したと見なされるタイプ」の欠落エラーが若干数残存するのみであり、中国審決コーパスデータの不足が見られるケースは存在しない。他方、「過去学習回では訳出できていた用語が突発的に欠落するタイプ」の欠落エラーも大規模アップデート以降確実に改善されており、現「みんなの自動翻訳」において欠落エラーが多発する可能性はきわめて小さいといえる。

4.5.5.7. 「湧き出しエラー」のカウント結果

「湧き出しエラー」とは、原文中に存在しない不要な語や表現が翻訳文中に含まれているエラーを指す。原文に一度しか出現しない語が翻訳文で複数回出現するような場合も、不要な2回目以降は湧き出しエラーとなる。ただし、内容理解を助ける／妨げない適切な補足（例えば第3.4.2.項の実例3-2における「以下のとおりである」の補足など）は湧き出しとは見なさない。

湧き出しエラーは、従来のルールベース方式による機械翻訳システムでは方法論的にほぼ発生せず、ニューラル翻訳など学習データを解析して翻訳文を生成する方式の機械翻訳に特徴的なエラーである。まずはカウント結果を下表に示す。

表 4.5.12. 各回学習前後の「湧き出し」該当文数カウント結果

エンジン	無学習	第5回学習	第10回学習	第15回学習	第20回学習
汎用 NT	14	3	6	0	0
	—	-11	+3	-6	±0
特許 NT	30	9	8	0	0
	—	-21	-1	-8	±0

両エンジンとも、中国審決コーパスデータの初回学習を経た第5回学習後にエラーが大きく減少し、第10回後には停滞したが第15回後に再び大きく減少し、以降は一切発生しなくなった。この推移は、無学習時のエラー発生数が特許 NT で顕著に多かったことを除けば、前項に示した欠落エラーとおおむね同様である。

まず、無学習時において特許 NT でのエラー発生数が汎用 NT よりも顕著に多かった点について考察する。結論から述べると、これは特許 NT が特許文献由来の対訳コーパスのみを基礎学習に使用していることによる副作用である可能性が高い。以下、実例で示す。

【文番号：#122、特許 NT】

原文	可见，对比文件 4 给出了如何降低基础表达的技术手段，且作用与其在本申请中为解决发明实际解决的技术问题所起的作用相同。
正解訳文	このように，引用文献 4 には基礎発現をいかにして低減するかを技術的手段が示されており，その作用は本願の発明が実際に解決しようとする技術的課題を解決するための作用と同じである。
無学習 内容伝達 Lv：1	図から分かるように，参考文献 4 は基礎発現の技術的手段をどのように低下させるかを示し，且つ本出願で発明の実際の解決を解決する技術的課題を解決するために作用する役割と同じである。

	－	BLEU : 0.2735 (－) RIBES : 0.8299 (－)
第 5 回学習後 内容伝達 Lv : 5		このことから分かるように,引用文献 4 は如何にして基本発現を低減するかという技術手段を与えており,且つその作用は本願において実際に解決しようとする技術的課題を解決するために果たす作用と同じである。
	向上	BLEU : 0.5100 (+0.2365) RIBES : 0.9278 (+0.0979)

上掲のとおり、特許 NT の無学習時の機械翻訳文では冒頭部に湧き出しエラーが発生している。原文冒頭の「**可见**」は中国審決では頻出の、前段の内容を受けて後続に接続する語であり、ニュアンスとしては正解訳の「**このように**」や第 5 回学習後の翻訳文の「**このことから分かるように**」等が適切であるが、これに対し特許 NT の無学習時の翻訳文は「**図から分かるように**」と、原文では全く言及されていない「**図**」という要素が湧き出ししている。なお、汎用 NT ではこの湧き出しエラーは発生していない⁴³。もう一例、同種の実例を示す。

【文番号：#539、特許 NT】

原文	驳回決定所依据的审查文本为：申请人于 2012 年 09 月 05 日提交的说明书第 1-274 段、说明书附图、说明书摘要、摘要附图，及 2016 年 04 月 11 日提交的 权利要求第 1-5 项 。	
正解訳文	拒絶査定の対象となった出願書類は以下のとおりである：出願人が 2012 年 9 月 5 日に提出した明細書第 1～274 段落、明細書の添付図面、要約書、要約書の添付図面，及び、2016 年 4 月 11 日に提出した 請求項第 1～5 項 。	
無学習 内容伝達 Lv : 1	拒絶決定に基づく審査テキストは以下のとおりである：出願人は 2012 年 9 月 05 日に提出された明細書第 1-274 段落、明細書図面、要約書、要約図面及び 2016 年 4 月 11 日出願の 特許請求の範囲第 1-5 項を参照されたい 。	
	－	BLEU : 0.3788 (－) RIBES : 0.8715 (－)
第 5 回学習後 内容伝達 Lv : 5	拒絶査定の対象となった出願書類は以下のとおりである：出願人が 2012 年 9 月 05 日に提出した明細書第 1-274 段落、明細書添付図面、要約書、要約書添付図面,及び 2016 年 4 月 11 日に提出した 請求項第 1-5 項 。	
	向上	BLEU : 0.7727 (+0.3939) RIBES : 0.9842 (+0.1127)

⁴³ 本例の汎用 NT の翻訳文は、技術／一般誤訳のカウント結果の項 (⇒4.5.5.2.) で「基礎発現」の誤訳事例 (P.140～) として使用しているので参照されたい。

本例では、原文文末の「**权利要求第 1-5 項**」が、正しくは「**請求項第 1～5 項**」と体言止めであるべきところ、特許 NT の無学習時の翻訳文では「**特許請求の範囲第 1-5 項を参照されたい**」と、原文に存在せず、文脈ともそぐわない「**を参照されたい**」という語句が湧き出している。なお、このエラーも汎用 NT では発生していない。

これら二例に代表されるように、無学習時の特許 NT では、特許文書において原文中の用語と組み合わせて使われがちな語が、原文の内容にかかわらず湧き出すケースが散見された。つまり「**分かるように**」であれば「**図**」、「**請求項**」であれば「**参照されたい**」など、特許文書で頻出の言い回しが原文と無関係に、連想的に出力されている。学習データに基づいて蓋然性の高い訳語を原文にとらわれず選定するという手法自体は、学習データを用いる方式の機械翻訳のセールスポイントというべきものであるが、無学習時の特許 NT では、上記二例のように、特許文献に由来する対訳コーパスを重点的に学習させたことの副作用と見られる湧き出しエラーもいくつか検出された。具体的には上記二例に加え文番号#1（「**発明**」であるべきところ「**本発明**」）、#1328（「**請求項 1**」であるところ「**請求項 1 に記載の発明**」）、#1606、#1651、#1893 などが挙げられる。これが、無学習時から初期学習時にかけて特許 NT の湧き出しエラーのカウントが多かった理由であり、この期間の内容伝達レベルが汎用 NT よりも劣っていた一因でもあるであろう。

とはいえ、上記二例の湧き出しエラーは上掲のとおりいずれも第 5 回学習後には解消されており、第 10 回学習後以降はカウント上も両エンジン間の差はなくなった。先に第 3.4.1.項や第 4.5.4.3.項において、基本的な特許用語への対応度に優るという特許 NT のアドバンテージは中国審決コーパスデータの学習により早々に失われたことを指摘したが、今回の結果から、それと同時に副作用も解消されたことがわかる。したがって、中国審決コーパスデータを学習データとして適用する前提であれば、こうした特許文献に重点を置いた基礎学習による副作用を理由に特許 NT の使用を忌避する必要はない。

続いては、両エンジンに共通して見られるタイプの湧き出しエラーについて分析する。まずは実例を一つ示す。

【文番号：#1745、汎用 NT】

原文	对此，合议组经审查后认为：(1) 本专利说明书第 5 段记载了“专利文献 3 中所记载的电晕装置在抑制上述长期 进行 电晕处理时的结晶状异物附着这一方面未必一定得到充分满足”，可见，电晕处理时薄膜上存在结晶状异物是申请日前本领域已知的内容。
----	---

正解訳文	これに対し、合議体は審理を経て以下のとおり判断した：(1) 本特許の明細書第 5 段落には「特許文献 3 に記載のコロナ装置は上述した コロナ処理を長期にわたって行う際の 結晶状異物の付着を 抑制する という点で、必ずしも満足のものではなかった」と記載され、ここから、コロナ処理時にフィルム上に結晶状異物が存在することは出願日より前に本分野では既知の内容である。	
無学習 内容伝達 Lv : 2	これに対して、合議組は、(1) 本特許明細書の第 5 段に、「特許文献 3 に記載されているコロナ装置が、これらの 長期的なコロナ処理を抑制する際の 結晶状異物の付着を 抑制する という点で必ずしも十分に満足されるとは限らない」とし、コロナ処理時には、フィルム上に結晶状異物が存在することが、申請前の本分野で知られていることを確認した。	
	—	BLEU : 0.3694 (—) RIBES : 0.8324 (—)
第 5 回学習後 内容伝達 Lv : 4	これに対し、合議体は審理の結果、以下のように判断した：(1) 本特許の明細書第 5 段落には、「特許文献 3 に記載のコロナ装置は、上記の コロナ処理を行う際の 結晶状異物の付着を 抑制する という面では必ずしも十分に満足できるものではない」と記載されており、このことから分かるように、コロナ処理時にフィルム上に結晶状異物が存在することは、出願日より前の本分野で公知の内容である。	
	向上	BLEU : 0.6070 (+0.2376) RIBES : 0.9267 (+0.0943)

上掲は無学習時と第 5 回学習後の汎用 NT の機械訳文である。無学習時の訳文では、本来「**コロナ処理を長期にわたって行う際の**」という意味であるべき箇所が「**長期的なコロナ処理を抑制する際の**」となっており、「**行う**」と訳されるべき箇所に「**抑制する**」という全く異なる意味の語が出力されている。「**行う**」と「**抑制する**」では意味がかい離しており、単なる誤訳と見なすには無理がある。このため本件は、「**進行**」の欠落エラーと、その埋め合わせとしての「**抑制する**」の湧き出しエラーとが同時に発生したものと判断した。

大規模アップデート以前の「みんなの自動翻訳」では、本例のように欠落エラーとその欠落を埋めるための湧き出しエラーとが同時に発生するケースが数多く見られた。例えば第 3.4 項の実例 3-3 では第 2 回～第 4 回学習後に「破乳ステップ (解乳化ステップ)」という語が「**乳化重合ステップ**」と誤訳されていたが、これは「破乳」の「破」が解析できず訳語が欠落し、その埋め合わせとして原文中の別の位置にある「**聚合**」の訳語である「**重合**」が重複して出力されたものと見なされる。

「みんなの自動翻訳」で欠落の埋め合わせとして湧き出しエラーが発生する場合、このよ

うに文中に存在する何らかの語が重複して出力されることが多い。本例もエラー箇所の直後に「結晶状異物の付着を抑制する」という一節が存在しており、この「抑制する」が重複して欠落箇所を埋め合わせている。

本例における無学習時の湧き出しエラー（及びそれを誘発した欠落エラー）は、上掲のとおり第5回学習後には解消された⁴⁴。したがって、ここまで本項で取り上げた湧き出しエラー三事例はいずれも、無学習時に湧き出しエラーが発生したが、学習初期の時点で解消された形となる。各回のカウントに見られるとおり、全体を通してはこのような推移を辿るケースが最も多かったと考えられる。

ただし、必ずしも全ての湧き出しエラーが無学習時から発生していて、それが学習の進行につれ徐々に解消されたわけではない。むしろ、湧き出しエラーは各回において前後の回との脈絡なくランダム的に発生する傾向が強く、全体としては徐々に発生件数が少なくなっているとはいえ、個々のケースを見てみると、それまでエラーが発生していなかった文で突如発生するようなケースも多い。この傾向から、湧き出しエラーも一部の欠落エラーと同様、特定の学習回における構文解析結果の変化などによって突発的に発生する性質を有すると考えられる。このため、例えば汎用 NT の第10回学習後のように、学習の進捗と反比例してエラー発生数が増加してしまうことも起こりうる。

実際、本例においても、無学習時に発生した湧き出しエラーは第5回学習後には解消されたが、第10回学習後には新たな湧き出しエラーが発生している。以下、第10回学習後と、それ以降の各回の翻訳結果を示す。

第10回学習後 内容伝達Lv:3	これに対し、合議体は審理の結果、以下のように判断した:(1)本特許明細書第5段落には「特許文献3に記載のコロナ装置は上記 コロナ処理時のコロナ処理時の 結晶状異物付着を抑制するという点で必ずしも満たされるべきではない」と記載されており、このことから分かるように、コロナ処理時にフィルムに結晶状異物が存在することは出願日より前に本分野で知られている内容である。	
	劣化	BLEU : 0.5924 (-0.0146) RIBES : 0.9131 (-0.0136)
第15回学習後 内容伝達Lv:4	これに対し、合議体は審理を経て以下のように判断した:(1)本実用新案の明細書第5段落には「実用新案登録公報3に記載のコロナ装置は、 コロナ処理を長期間行った際 の上記結晶状異物の付着を抑制する点で	

⁴⁴ ただし、無学習時には正しく訳出されていた「**長期的な**」というニュアンスが第5回時の翻訳文では欠落している。

	必ずしも十分に満足できるものではない」と記載されており,以上から分かるように,コロナ処理時にフィルム上に結晶状異物が存在することは出願日より前に本分野で知られている内容である。	
	向上	BLEU : 0.5963 (-0.0039) RIBES : 0.9337 (+0.0206)
第 20 回学習後 内容伝達 Lv : 4	これに対し,合議体は審理の結果、以下のように判断した:(1)本実用新案の明細書第 5 段落には「特許文献 3 に記載されているコロナ装置は、 コロナ処理を長期間行う時の 上記結晶状異物の付着を抑制するという点で必ずしも十分に満たされるわけではない」と記載されており,このことから分かるように,コロナ処理時にフィルムに結晶状異物が存在することは出願日より前の本分野で既知の内容である。	
	劣化	BLEU : 0.5565 (-0.0398) RIBES : 0.9205 (-0.0132)

第 10 回学習後に発生した湧き出しエラーは、発生箇所は無学習時と同じだが内容が異なり、「**コロナ処理を長期にわたって行う際の**」であるべきところ「**コロナ処理時のコロナ処理時の**」と、「**長期にわたって行う**」が欠落し、その埋め合わせとして直前の「**コロナ処理時**」が重複して湧き出す形となっている。同一文の同じ箇所で繰り返しエラーが発生していることから、本例にはこの箇所の構文解析を失敗しやすい何らかの要素が含まれている可能性が高いが、それでも全回を通じてこの箇所に湧き出しエラーが発生したのは、汎用 NT では上掲の 2 回以外は第 2 回学習後のみ、特許 NT も第 1 回と第 11 回学習後のみであり、さほど頻発しているわけではない。

こうした発生状況は多くの湧き出しエラーに共通しており、かつ欠落エラーの第 2 のタイプに酷似する。つまり、湧き出しエラーも欠落エラーと同様、エンジンのロジックによって突発的に発生する性質のエラーであると考えられる。このため毎回一定量発生し、学習データを追加しても改善は難しい。事実、前掲のカウント結果でも第 5 回から第 10 回にかけてはエラーはほとんど減少していない⁴⁵。

これに対し、第 15 回学習後以降は湧き出しエラーが全く発生しなくなった。これも第 2 のタイプの欠落エラー同様、第 12 回学習前に実施された「みんなの自動翻訳」の大規模アップデートの効果と考えられる。これまでの分析から、本アップデート以降、自動評価スコア、内容伝達レベルともスコア水準が明らかに向上するとともに、それまで顕著であった一般/技術用語や文構造におけるランダム的な訳語変化が抑制され、学習データ中の最有力

⁴⁵ 無学習時から第 5 回学習後にかけては湧き出しエラー数が減少しているが、これは湧き出しエラーが欠落エラーの埋め合わせとして発生することが多く、トリガーとなる欠落エラーがこの時期に減少したためと考えられる。第 5 回学習後までに、審決用語が学習されていないことによる欠落エラー（第 1 のタイプ）が、コーパスの学習によって一挙に解消している。

の訳語候補が採用される確率が高まったことが確認されている。こうしたエンジンの根幹的ロジックの見直しの際に、同じく旧ロジックに起因していた湧き出しエラーの発生を根本的に解消する何らかの対策が講じられたものと推察される。例えば第 2 のタイプの欠落エラーや、文末に不要な「2」が出力されるエラー（実例 4-1 参照）などのエラーが同じタイミングで完全に解消されている。ロジックが改良された以上、今後「みんなの自動翻訳」において湧き出しエラーが再発する懸念はきわめて小さい。

4.5.5.8. 「数値エラー」のカウント結果

本カテゴリには、原文中の数字が翻訳文中で別の数字になっている、本来あるべき位置に配置されていないなど、原文中の数値に起因すると思われる不備全般をカウントした。数値自体は正しくても、数値の処理が影響している判断されるようなエラーであれば、それらも数値エラーと見なした。以下、カウント結果を示す。

表 4.5.13. 各回学習前後の「数値エラー」該当文数カウント結果

エンジン	無学習	第 5 回学習	第 10 回学習	第 15 回学習	第 20 回学習
汎用 NT	23	11	10	5	5
	—	-12	-1	-5	±0
特許 NT	22	12	10	5	5
	—	-10	-2	-5	±0

無学習時から学習初期にかけての数値エラーの代表例は、「原文中の数値が文末に羅列されるエラー」であった。以下、実例を示す。

【文番号：#704、汎用 NT】

原文	(3) 请求人当庭放弃新颖性的无效理由, 放弃证据 1 和证据 3 作为证据使用, 并明确其无效宣告请求的范围、理由以及证据使用方式为: 以证据 2 作为最接近的现有技术, 权利要求 1-3 相对于证据 2 和公知常识的结合不具备创造性, 权利要求 5 相对于证据 2、证据 4 以及公知常识的结合不具备创造性, 不符合专利法第 22 条第 3 款的规定。
正解訳文	(3) 請求人は当審において新規性に関する無効理由を放棄し、証拠 1 及び証拠 3 を証拠として使用することを放棄した上で、その無効宣告請求の範囲、理由及び証拠の使用方法を以下のとおり明確にした: 証拠 2 を最も近い従来技術とし、請求項 1～3 が証拠 2 及び技術常識の組み合わせと比較して創造性を有せず、請求項 5 が証拠 2、証拠 4 及び技術常識の組み合わせと比較して創造性を有せず、特許法第 22 条第 3 款の規定に適合しない。
無学習 内容伝達 Lv: 1	(3) 要求者が庭放に奇抜性を棄却する理由, 証拠を棄却する 1 と証拠 3 を証拠として使用する, その無効宣言要求の範囲, 理由および証拠使用方式: 証拠 2 を最も近い既存の手法として, 権利要求 1-3 を証拠 2 と公知の結合に対して創造性を持たない, 特許法第 22 条の第 3 項の規定に対しては, 特許法の第条の規定に反する. 5 2 4。
	—

第 5 回学習後 内容伝達 Lv : 5	(3)請求人は当審において新規性の無効理由を取り下げ,証拠 1 と証拠 3 を証拠として使用することを放棄し,且つその無効宣告請求の範囲、理由及び証拠の使用方法を以下のとおり明確にした:証拠 2 を最も近い従来技術とした場合,請求項 1-3 は証拠 2 と技術常識の組み合わせに対して創造性を有さず,請求項 5 は証拠 2、証拠 4 及び技術常識の組み合わせに対して創造性を有さず,特許法第 22 条第 3 款の規定に適合しない。	
	向上	BLEU : 0.6781 (+0.5045) RIBES : 0.9302 (+0.2185)

無学習時の翻訳文では、文末に「524」と算用数字が列挙されている。これらの数字は、正解訳文では文中の「…請求項 5 が証拠 2、証拠 4 及び…」という箇所が登場するが、無学習時の翻訳文は、そもそもこの箇所全体が欠落してしまっており、そこに含まれていた数値のみが文末に列挙された形である。

本例を始め多数の数値エラーの状況から総合的に判断して、「みんなの自動翻訳」では、原文中の数字部分を前処理で一旦ワイルドカードに置換し、その後、その状態で翻訳文を生成したのちに、エンジン内に記憶しておいた各数字を改めてワイルドカードと置き換える、という処理を行っているものと推察される。その際、本例のように翻訳文の生成に失敗して文の一部がワイルドカードごと消失してしまうと、記憶していた数字が行き場を失い、やむなく文末に列挙されるようである。この「原文中の数値が文末に羅列されるエラー」は、第 3.4.6.項に例示した、原文が長大すぎて大規模な欠落が発生している文でも多数発生していた。

したがってこのエラーは、本質的には数値の処理の問題ではなく、大規模な欠落エラーから派生したものといえる。本例においても、第 5 回学習後の機械翻訳結果では欠落エラーが解消した結果、各数字も文中の正しい位置に出力されるようになった。

ただし、このように文末に数字が出力される数値エラーの中には、欠落エラーとは関係なく、単独で発生するタイプも存在した。それらには、原文中に存在しない算用数字 2 が文末に出力されるという共通点があった。この数値エラーは人手評価対象 100 文中では第 4.5.4.1.項で取り上げた実例 4-1 (文番号#1494) のみで発生していたが、評価用コーパス全体では他にも多数発生している。以下、一例を示す。

【文番号 : #1306、汎用 NT】

原文	此竖直板 (13) 的左右两端的转角 (13A) 处, 突出部分 (10) 和安装部 (12) 同向并且如图 2 所示设置成 L 型弯曲状。
----	--

正解訳文	この立壁板(13)の左右両端の角部(13A)には、突出部分(10)が取付け部(12)と同方向に向いて且つ図2に示すとおりL型の湾曲状に設置される。	
無学習 内容伝達Lv: -	前記垂直板(13)の左右両端の回転角(13A)は、突出部分(10)と取付け部(12)が同方向且つ図2に示すようにL型湾曲状に設置される。2。	
	-	BLEU: 0.4800 (-) RIBES: 0.9311 (-)
第5回学習後 内容伝達Lv: -	この垂直板(13)の左右両端のコーナー(13A)において、突出部(10)と取付け部(12)は同方向であり、且つ図2に示すようにL字状に湾曲するように設置される。2。	
	-	BLEU: 0.4049 (-0.0751) RIBES: 0.9033 (-0.0278)
第10回学習後 内容伝達Lv: -	この垂直板(13)の左右両端のコーナー(13A)において、突出部分(10)及び取り付け部(12)は同方向に配置され、且つ図2に示すようにL字状に設けられる。2	
	-	BLEU: 0.4217 (+0.0168) RIBES: 0.8962 (-0.0071)

※人手評価対象文ではないため内容伝達レベルや学習前後の相対評価は実施していない。

本例も実例4-1と同様、無学習時から第10回学習後にかけて、翻訳文の文末に常時「2。」又は「2」が出力されている。原文中に「図2」や「安裝部(12)」として算用数字2が存在するので紛らわしいが、これらは翻訳文でも「図2」「取付け部(12)」とそれぞれ正しい位置に出力されており、文末に出力された「2」の出所ではない。これは、原文中に算用数字2を含まない多くの事例、例えば文番号#198、#610、#1621などでも同じエラーが発生していることから明らかである。

前例の分析において、翻訳文末尾に羅列された数字が原文中に存在した数字が前処理時にワイルドカードに置換されエンジン内に一時記憶されたが、欠落エラーによりワイルドカードが消失して行き場を失った結果、やむなく文末に出力されたものと説明した。このメカニズムに従えば、本例を始め文末に2が出力されるケースでは、原文中に存在しない数値2がエンジン内に記憶されていたことになる。

この観点で本例の原文を見ると、数値2に準ずる漢字として、日本語の「両」にあたる「兩」が存在していることに気づく。そして本エラーが発生している各原文の全てに、漢数字の「二」か、若しくは本例の「兩」のように「2」のニュアンスを有する漢字が、語の一部として存在していることが判明した。「語の一部として存在」とは、本例で言えば「兩」という文字が「両端(両端)」、文番号#198、#610では「兩者(両者)」、そして#1621では「二」が「聚丁二醇(ポリテトラメチレングリコール)」という用語の一部として出現していることを指す。

こうした状況から、当時の「みんなの自動翻訳」では、漢数字やこれに準ずる「兩」などの漢字に対しても、これを数字に読み替えてエンジン内に「2」が記憶されているものと推察される。だが、「兩」は数値ではなく「両端」という語句の一部であるため、翻訳処理では「両端」とワイルドカードを含まない形で訳される。その結果、エンジン内に記憶された「2」は行き場を失い、文末に出力されたものと考えられる。例えば原文に「兩者」を含む文番号#649では、この語が「2者」と訳された無学習時には文末に「2」は出力されておらず、「兩者」と訳されるようになった第1回学習後から文末に「2」が出力されている。

つまり、本エラーは突発的に発生するエラーではなく、原文中に漢数字「二」若しくはそれに準ずる漢字が語の一部として存在し、その語が語として正しい単位で扱われた場合には確実に発生する性質のエラーである。本例においても、両エンジン、無学習時から第10回学習後にかけて基本的には毎回発生している⁴⁶。

本エラーは、エンジン内の数値の取り扱いに起因するものであって、学習データによる改善は不可能である。だが、「みんなの自動翻訳」の大規模アップデートを経た第12回学習後以降、本エラーは一切発生しなくなった。以下、本例の第15回学習後及び第20回学習後の機械翻訳結果を示す。

第15回学習後 内容伝達Lv：－	この垂直板(13)の左右両端の回転角(13A)において、突出部分(10)と取り付け部(12)は同じ方向であり且つ図2に示すようにL字型湾曲状に設置される。	
	－	BLEU：0.5070 (+0.0853) RIBES：0.9278 (+0.0316)
第20回学習後 内容伝達Lv：－	この垂直板(13)の左右両端のコーナ(13A)において、突出部分(10)と取り付け部(12)は同方向であり、且つ図2に示すようにL字型湾曲状に設置される。	
	－	BLEU：0.5114 (+0.0044) RIBES：0.9278 (±0)

前項、前々項で述べたとおり、大規模アップデートを境に、エンジンのロジックに起因するタイプの欠落エラーや湧き出しエラーが全面的に解消されたが、本例における改善状況から、上記「2」の数値エラーを起こしていたロジックも同時に見直されたものと考え

⁴⁶ 特許NTの無学習時、第4回学習後、第8回学習後のみは、文中の「转角(13A)」が「回転角(2A)」と訳され、文末には13が出力されるという数値の入れ替わりが発生している。これも、ワイルドカードに一次置換した数値を元に戻す際、エンジン内に記憶された各数値から誤ったものを戻してしまったためと考えられる。

られる。大規模アップデート以降、それまで多数の文⁴⁷で毎回発生していた同種のエラーは完全に解消され、その結果、数値エラーの総カウント数も半減した。

ただし、大規模アップデート以降も二種類の数値エラーが最終学習後まで残存している。一つは「日付における前ゼロ残留エラー」である。以下、実例を示す。

【文番号：#1078、特許 NT】

原文	因此，本案的审查基础是专利权人在 2017 年 03 月 06 日提交的权利要求第 1-6 项。	
正解訳文	これにより，本案の審理は特許権者が 2017 年 3 月 6 日に提出した請求項第 1 - 6 項を基礎とする。	
無学習 内容伝達 Lv : 2	したがって、本発明の審査基礎は、2017 年 3 月 06 日に出願された特許請求の範囲の請求項 1-6 である。	
	—	BLEU : 0.0768 (—) RIBES : 0.6244 (—)
第 5 回学習後 内容伝達 Lv : 4	したがって、本案の審理の基礎は実用新案権者が 2017 年 3 月 06 日に提出した請求項第 1-6 項である。	
	向上	BLEU : 0.4979 (+0.4211) RIBES : 0.8699 (+0.2455)
第 10 回学習後 内容伝達 Lv : 4	従って、本案の審理の基礎は実用新案権者が 2017 年 3 月 06 日に提出した請求項 1-6 である。	
	同等	BLEU : 0.4215 (-0.0764) RIBES : 0.8437 (-0.0262)
第 15 回学習後 内容伝達 Lv : 4	従って、本案の審理の基礎は実用新案権者が 2017 年 3 月 06 日に提出した請求項 1-6 である。	
	同等	BLEU : 0.4215 (±0) RIBES : 0.8437 (±0)
第 20 回学習後 内容伝達 Lv : 4	従って、本案の審理の基礎は実用新案権者が 2017 年 3 月 06 日に提出した請求項 1-6 である。	
	同等	BLEU : 0.4215 (±0) RIBES : 0.8437 (±0)

本例の原文では、日付は「2017 年 03 月 06 日」と、月・日とも前ゼロありとなっている。本例を含め、中国審決原文では日付は原則としてこの「前ゼロあり」フォーマットで統一されている。一方、本事業では日付の前ゼロは一律除去して翻訳する方針とした。このため正解訳文も「2017 年 3 月 6 日」と、月・日とも前ゼロが除去されている。

⁴⁷ 評価用コーパス 2,013 文のうち、汎用 NT の無学習時の翻訳文の文末が「2。」又は「2」であった文は 54 文存在した。

これに対し、「みんなの自動翻訳」では、本例各回の翻訳結果（2017年3月06日）に見られるとおり、「月の場合は前ゼロ除去、日の場合は前ゼロのまま」という中途半端なフォーマットが採用されている。原文日付の日部分が前ゼロ⁴⁸である全ての日付において本例と同様のエラーが常時発生しており、大規模アップデート後も改善は見られなかった。

この「月は前ゼロ除去、日は前ゼロ維持」というフォーマットは中国審決コーパスデータから学習されたものではありえず、「みんなの自動翻訳」自体の設定で強制的に採用されていると考えられる。このため、改善にはエンジン自体の設定変更が必要である。内容理解を阻害しないごく軽微な不備であり、最優先で取り組むべき課題ではないが、エンジン開発元であれば設定変更により容易に対処可能と考えられる。

最終学習後に残存したもう一種の数値エラーは、より重大である。以下、実例を示す。上例で取り上げた文番号#1078の、汎用 NT での機械翻訳文である。

【文番号：#1078、汎用 NT】

原文	因此，本案的审查基础是专利权人在 2017 年 03 月 06 日提交的权利要求第 1-6 项。	
正解訳文	これにより，本案の審理は特許権者が 2017 年 3 月 6 日に提出した請求項第 1 - 6 項を基礎とする。	
無学習 内容伝達 Lv : 3	そのため，本案の審査基礎は，特許権者が 2017 年 3 月 06 日に提出した請求の第 1-6 項である。	
	—	BLEU : 0.3872 (—) RIBES : 0.8396 (—)
第 5 回学習後 内容伝達 Lv : 5	したがって，本案の審理の基礎は特許権者が 2017 年 3 月 06 日に提出した請求項 1-6 である。	
	向上	BLEU : 0.4853 (+0.0981) RIBES : 0.8597 (+0.0201)
第 10 回学習後 内容伝達 Lv : 5	したがって，本案の審理の基礎は特許権者が 2017 年 3 月 06 日に提出した請求項 1-6 である。	
	同等	BLEU : 0.4853 (±0) RIBES : 0.8597 (±0)
第 15 回学習後 内容伝達 Lv : 5	したがって，本案の審理の基礎は特許権者が 2017 年 3 月 06 日に提出した請求項 1-6 である。	
	同等	BLEU : 0.4215 (-0.0638) RIBES : 0.8437 (-0.0160)
第 20 回学習後 内容伝達 Lv : 4	従って，本案の審理の基礎は特許権者が 2017 年 3 月 1 日に提出した請求項 1-6 である。06	

⁴⁸ つまり、日が 01 日～09 日である全ての日付が該当する。

	劣化	BLEU : 0.4858 (+0.0643) RIBES : 0.8437 (±0)
--	----	---

本例は汎用 NT においても、無学習時から第 15 回学習後にかけて特許 NT と同じく「日付における前ゼロ残留エラー」が発生しているが、ここでは論点としない。問題は第 20 回時の機械翻訳文において、日付部分が「2017 年 3 月 1 日」と、原文と異なる日付になってしまっている点である。なお、本来の日付である「06」は文末に数値として別途出力されている。

同様のエラーは人手評価対象 100 文ではもう一文、文番号#539 でも発生しており、こちらは原文の「09 月 05 日」が「9 月 1 日」と訳され、文末に「05」が出力されている。この【日付が 1 日に変化する】という新規エラーはこれら 2 文とも汎用 NT のみで発生している。

さらに詳しく調査した結果、これら 2 文とも本エラーは第 19 回学習後から発生していた。また、評価用コーパス全件においては、第 20 回学習後の時点で文番号#5、#7、#113、#170、#174 など全 1,994 文のうち 46 文において同様のエラーが発生していることも判明した。全て汎用 NT のみの発生であったが、文によって第 19 回後から発生しているものと第 20 回後のみ発生しているものが混在する。

特許 NT では評価用コーパス全件に範囲を広げても同エラーは一切検出されず、このことから本エラーは汎用 NT 特有のものである可能性が高い。ただし、日付を含む全ての文でエラーが発生しているわけではなく、発生条件は不明である。また、大抵の場合は上例と同じく本来の日付が「1 日」に置き換わっているが、文番号#1517 では原文の「07 月 07 日」が翻訳文では「7 月 8 日」に変化するというイレギュラーなケースも見られた。このため本エラーは正確には「日付の日部分が異なる日付に変化するエラー」と呼ぶべきであろう。

本エラーも「日付における前ゼロ残留エラー」と同様、明らかにエンジンのロジックに起因している。かつ、記載内容の誤解を招くという点で「日付における前ゼロ残留エラー」よりも重大なエラーであり、開発元による対処が必須である。

4.5.5.9. 「記号エラー」のカウント結果

本カテゴリには、ピリオドやコロン、カッコやクォーテーションマークなど、各種記号の取り扱いに関連する不備をカウントした。以下に結果を示す。

表 4.5.14. 各回学習前後の「記号エラー」該当文数カウント結果

エンジン	無学習	第 5 回学習	第 10 回学習	第 15 回学習	第 20 回学習
汎用 NT	8	2	1	0	0
	—	-6	-1	-1	±0
特許 NT	8	1	1	0	0
	—	-7	±0	-1	±0

本カテゴリに該当するエラーは無学習の時点から少数であり、かつその内容も「引用箇所がカッコで括られていない (#351、特許 NT)」、「文末が句点でなく半角ピリオドになっている (#462β、汎用 NT)」、「『請求項 1-3〜』と余計な「〜」が存在する (#1010、特許 NT)」など、内容伝達への悪影響の少ない軽微な不備のみであった。かつ、その大半は学習初期に解消されており、第 15 回学習後以降は両エンジンとも一切発生しなくなった。

4.5.5.10. 「その他エラー」

上記各カテゴリに該当しない不備は全てここに分類した。以下、各回のカウント結果を示す。

表 4.5.15. 各回学習前後の「その他エラー」該当文数カウント結果

エンジン	無学習	第 5 回学習	第 10 回学習	第 15 回学習	第 20 回学習
汎用 NT	2	0	0	0	0
	—	-2	±0	±0	±0
特許 NT	4	2	1	0	0
	—	-2	-1	-1	±0

カウントは上表のとおり無学習時からごく少数であり、このことから、本調査分析で設定した前出 9 種のカテゴリで、機械翻訳で生じるエラーの大部分がカバーされていたことがわかる。本カテゴリにカウントされた事例はいずれも「訳ゆれ」、すなわち「一文中に複数回出現する用語が異なる訳語に訳されている（その結果、内容理解を阻害する）」という不備の指摘であった。以下、典型例を示す。

【文番号：#1146、特許 NT】

原文	<p>(2) 对比文件 3 与权利要求 1 的安装结构不同, 对比文件 3 电机壳体通过膏状舒焊涂敷能够使平坦状底面部粘着在印刷配线板的固定用图形上进行固定, 是通过电机壳体直接将电机固定在印刷配线板上, 没有相应的独立于电机壳体之外的安装件, 而本专利权利要求 1 中是通过电机主体以外的安装件固定电机, 因此对比文件 3 没有公开安装件以及定位凹部、插入支持片和底部提高平板状部能够收纳在所述马达壳体的粗度以内”。</p>
正解訳文	<p>(2) 引用文献 3 と請求項 1 の取付け構造は異なり, 引用文献 3 のモータケースはペースト状半田コーティングによって平坦状底面部をプリント基板の固定用パターン上に接着固定して, モータケースによって直接モータをプリント基板上に固定するものであり, 相応のモータケースの外の独立した据付金具はなく, 更に本特許請求項 1 はモータ本体以外の据付金具を介してモータを固定するものであり, よって引用文献 3 は据付金具及び位置決め凹部、差し込み支持片及び底上げ平板状部が前記モータケースの太さ以内に収納できることを開示していない。</p>
無学習 内容伝達 Lv : 1	<p>(2)比較文献 3 と請求項 1 の取り付け構造が異なり,比較文献 3 のモータハウジングはペースト状舒焊塗布によって平坦状底面部をプリント配線板の固定用パターンに接着して固定することができ,モータハウジングによってモータを介してプリント配線板に直接固定され,モータハウジング以外の取付部材を介して固定され,モータハウジングの太さ以内「内」に収納することができる。13。</p>
	<p>— BLEU : 0.1804 (-) RIBES : 0.7814 (-)</p>
第 5 回学習後 内容伝達 Lv : 3	<p>(2)引用文献 3 は請求項 1 の取り付け構造とは異なり,引用文献 3 のモータケーシングはペースト状舒焊塗布によって平坦状底面部をプリント配線板の固定パターンに接着して固定することができ,モータケースを介してモータをプリント配線板に直接固定することであり,モータケーシングから独立した取り付け部材がなく,本特許の請求項 1 はモータ本体以外の取り付け部材によってモータを固定するものであり,従って引用文献 3 は取り付け部材及び位置決め凹部、挿入支持片及び底上げ平板部が前記モータケーシングの粗さ内に収納されることを開示していない」。</p>
	<p>向上 BLEU : 0.4426 (+0.2622) RIBES : 0.8744 (+0.0930)</p>
第 10 回学習後 内容伝達 Lv : 2	<p>(2)引用文献 3 と請求項 1 の取り付け構造が異なり,引用文献 3 のモータハウジングはペースト状舒焊を介して平坦状底面をプリント配線板の固定用パターンに塗布して固定することができ,モータケースを介し</p>

	てモータをプリント配線板に直接固定することは、 モータケース 以外の取り付け部材から独立したのではなく、本特許の請求項1ではモータ本体以外の取り付け部材によってモータを固定するため、引用文献3には取り付け部材及び位置決め凹部、支持板及び底板部を取り付けることが前記 モータハウジング の粗さ内に収容できることが開示されていない」。
	劣化 BLEU : 0.3141 (-0.1285) RIBES : 0.7923 (-0.0821)
第15回学習後 内容伝達Lv:3	(2)引用文献3は請求項1の取付構造とは異なり、引用文献3の モータケース はペースト状の舒焊を塗布することにより平坦状底面部をプリント配線板の固定用パターンに固着させて固定することができ、 モータケース によりモータをプリント配線板に直接固定し、対応する モータケース から独立した取付具がないが、本実用新案の請求項1ではモータ本体以外の取付具によりモータを固定するため、引用文献3には取付具及び位置決め凹部、挿入支持片及び底上げ平板部が前記 モータケース の太さ以内に収納できることが開示されていない」。
	向上 BLEU : 0.4157 (+0.1016) RIBES : 0.8414 (+0.0491)
第20回学習後 内容伝達Lv:4	(2)引用文献3は請求項1の取り付け構造と異なり、引用文献3の モータハウジング はクリームはんだ塗布により平坦状底面部をプリント配線板の固定用パターンに接着して固定することができ、 モータハウジング によりモータをプリント配線板に直接固定し、対応する モータハウジング から独立した取り付け具がないが、本実用新案の請求項1ではモータ本体以外の取り付け具によりモータを固定するため、引用文献3には取り付け具及び位置決め凹部、挿入支持片及び底部向上平板状部が前記 モータハウジング の太さ以内に収納できることが開示されていない。」
	向上 BLEU : 0.3679 (-0.0478) RIBES : 0.8332 (-0.0082)

本例は、正解訳文のとおり「**モータケース**」という名称が文中で4回繰り返されている。全て同一の構成要素を指しており、こうした語に訳ゆれが生じると、読み手の内容理解を大きく阻害する懸念がある。

各回の翻訳結果を見ていくと、無学習時は4回全て「**モータハウジング**」と、正解訳文とは違うものの同じ意味の語で統一されており、訳ゆれは生じていない。これに対し、第5回学習後は「**モータケーシング** (×3)」と「**モータケース** (×1)」とで訳ゆれが生じ、続く第10回学習後も、今度は「**モータハウジング** (×2)」と「**モータケーシング** (×2)」との間で訳ゆれした。

本例は原文自体、最初の三回は「**电机壳体**」、最後の一回は「**马达壳体**」と名称ゆれが生じているのだが、各回の訳ゆれ状況を見ると、原文で「**电机／马达**」と不統一な「モータ」ではなく、統一されている「**壳体**」の部分が「ケース」「ハウジング」「ケーシング」と訳ゆれしており、そもそも原文が統一されている三回の「**电机壳体**」間でも訳ゆれが生じているので、原文の問題とはいえない。かつ、各回で採用される訳語が目まぐるしく変化していることから、この訳ゆれには「みんなの自動翻訳」の特徴である技術／一般用語における訳語選定の不安定性が大きく影響している可能性が高い。これまで、こうした訳語の不安定な変化は同一文における学習前後での変化という見地から考察してきたが、本エラーにより、少なくとも大規模アップデート以前においては、同一学習回においても同じ用語の訳語がランダムに変化する可能性があることが判明した。

本例に関しては、大規模アップデート以降の第15回及び第20回学習後は、第15回後は「**モータケース**」×4、第20回後は「**モータハウジング**」×4と、訳語は違うものの、同一回での訳ゆれは発生していない。双方のエンジン、人手翻訳対象外の各回まで範囲を広げても、アップデート以前の各回のような同一文中での訳ゆれは発生しなくなった。とはいえ、この一例のみをもって、大規模アップデート以降の「みんなの自動翻訳」には同一文中の技術用語の訳ゆれの懸念は無くなったとは確言できない。また、技術用語は一文中のみで統一されているだけでなく、少なくとも同一案件上の各文において統一されていないと、読み手に無用の混乱を与える。こちらについては、ここまでの評価結果からは発生状況は把握できない。

本調査分析では、技術／一般用語など学習データ中に有力な訳語候補が複数存在するような語においては、大規模アップデート以降も学習前後で頻繁な訳語変化が生じていることが判明している。ニューラル機械翻訳における訳語選定のメカニズム(⇒3.4.3.)を考えると、こうした不安定な訳語変化が同一学習回の異なる文の間でも発生している懸念はむしろ大きい。この懸念に関しては、第5.4.項で改めて考察する。

5. 中国審決への機械翻訳の適用に関する分析・考察

本章では、第 3 章に示した自動評価結果及び第 4 章に示した人手評価結果の総合的な分析により得られた、「みんなの自動翻訳」及び「中国審決コーパスデータ」を用いた中国審決の機械翻訳文の提供に関する主な考察内容についてまとめる。

5.1. 全体的な品質改善効果について

本事業は、報告書冒頭の「1.1.本事業の背景・目的」に記したとおり、中国審決は専門性が高く、デフォルト状態のニューラル機械翻訳では十分な精度の翻訳結果は得られないとの前提のもと、中国審決由来の対訳コーパスを大量に作成し、これを機械翻訳に学習させることで品質改善を図ることを目的に実施された。

この観点から各評価結果を考察すると、まず評価用コーパス全件を対象とした自動評価においては、汎用 NT では BLEU の全体平均スコアが無学習時の 0.2874 から最終学習後は 0.6778 (+0.3904) へ、RIBES が 0.8013 から 0.9371 (+0.1358) へと、両評価指標とも大きく向上した。一方の特許 NT も BLEU が 0.3550 から 0.6748 (+0.3198) へ、RIBES が 0.8226 から 0.9352 (+0.1126) へと両指標で大幅な向上を果たしており、両エンジンともに、中国審決コーパスデータの学習によって中国審決文に由来する評価用コーパス全体の機械翻訳品質が顕著に改善されたことが示された。

評価用コーパス中の 100 文を対象に実施した人手評価においても、機械翻訳文の実際の品質を目視で判定した「内容伝達レベル」の評価（5 段階評価）では、100 文全件の平均レベルが汎用 NT では無学習時の 2.41 から最終学習後には 4.65 (+2.24) へ、特許 NT では 2.18 から 4.61 (+2.43) へと、やはり両エンジンとも 2 レベル強の大幅な品質改善を示しており、自動評価結果の妥当性が裏付けられた。

自動評価と人手評価の双方の結果が整合したことで、人手評価結果から看取される傾向は、評価用コーパス全体に押し広げて論じられる。つまり内容伝達レベルの評価結果から、中国審決コーパスデータの学習による具体的な品質改善の度合いが端的に把握できる。

内容伝達レベルの定義では、レベル 2 が「いくつかの重要情報は正確に伝達されている (20%~)」、レベル 3 が「半分以上の重要情報は正確に伝達されている (50%~)」、レベル 4 が「ほとんどの重要情報は正確に伝達されている (80%~)」にそれぞれ相当する(⇒4.5.1.)。この定義に照らせば、デフォルト時の「みんなの自動翻訳」による中国審決文の翻訳品質が概して「重要情報の 20%~50%程度が正確に伝達される」程度であったのに対し、中国審決コーパスデータを全件学習させることで、その翻訳品質は「重要情報の 80%以上が正確

に伝達される」レベルに改善されたことになる。

さらに、評価対象 100 文の文ごとの内容伝達レベルの分布を見ると、無学習時は両エンジンとも全体の 80%強の文がレベル 1~3 という低いレベルに属していたのに対し、最終学習後には 97%の文にスコア 4~5 が付されている (⇒4.5.1.)。コーパスによる品質改善効果が、頻出の定型文など一部の特定の文のみでなく、中国審決由来の多種多様な文に網羅的に奏されていることを示す結果である。

これらの結果から、本事業の前提とした、中国審決はデフォルト状態のニューラル機械翻訳では十分な精度の翻訳結果は得られないこと、及び本事業の目的である中国審決由来の対訳コーパスを大量に学習させることで品質改善が果たされることの双方が実証された。より具体的には、本事業で作成した中国審決コーパスデータ全件を「みんなの自動翻訳」に学習させることにより、中国審決の機械翻訳品質を、重要情報の 20~50%程度しか伝達されないレベルから、重要情報の 80%以上が正確に伝達されるレベルまで改善できることが証明された⁴⁹。

5.2. 特許・審決用語の訳質改善について

中国審決コーパスデータを学習させる主目的は、機械翻訳エンジンを中国審決特有の用語や表現に対応させることである。この目的がどの程度達成されたかは、人手評価における「エラーのカテゴリ別カウント」の「特許誤訳」のカウント推移により把握できる。

特許誤訳では、評価対象 100 文について、文中の特許・審決用語及び表現に誤訳があるか否かを目視で判定した。カウント結果は第 4.5.5.1.項に示したとおりであり、全 100 文中、無学習の時点では汎用 NT で 91 文、特許 NT で 90 文と、大多数の文で何らかの特許誤訳が発生していた。カテゴリ別にカウントした 10 種のエラーの中でも突出した発生頻度であり、デフォルト時の「みんなの自動翻訳」が、中国審決に用いられる特許・審決用語にほとんど対応できていなかったことがわかる。

これに対し、中国審決コーパスデータ全件を学習した第 20 回学習後における特許誤訳の発生は汎用 NT が 9 文、特許 NT が 16 文と、両エンジンとも大幅な改善を見せた。最終学習後まで残存した特許誤訳は全て「専利」の訳し分け誤りであり、この課題については第 5.5.項で後述するが、この一種類を除けば、中国審決文における特許・審決用語の誤訳・不適訳は、中国審決コーパスデータを学習させたことによって完全に解消されたといえる。

⁴⁹ なお、無学習時と第 20 回学習後とでは「みんなの自動翻訳」自体のバージョンが異なり、品質改善にはこうしたエンジン自体の性能向上が寄与した部分も含まれている。

自動評価結果、人手評価結果それぞれの分析で定点観測した各実例においても、ほぼ全ての特許・審決用語が学習初期から中国審決コーパスデータで使用されている所定の適訳語に改善され、以降の各学習回後も常時その適訳語が安定して採用され続けた。当初は正しく訳せなかった特許・審決用語がコーパスの学習によりことごとく適訳語で訳されるようになったことが、前項で述べた翻訳品質改善の最大の要因であることは明白であり、本事業で作成した中国審決コーパスデータが、機械翻訳システムを中国審決特有の用語・表現に適応させるという主目的を十分に果たすものであることが確認された。

5.3. 技術用語・一般用語の網羅性について

審決は発明や先行技術の内容にも言及するため、技術／一般用語も多数含んでいる。このため技術用語や一般用語の翻訳精度も中国審決の翻訳品質に影響する。

中国審決コーパスデータの主目的は特許・審決用語の学習であり、技術／一般用語を広く網羅しているわけではない。したがって、技術／一般用語の翻訳品質は、基本的には機械翻訳システム自身の基礎学習によるデフォルトの翻訳精度に依存する。ただし、中国審決コーパスデータ自体も請求項など原出願の摘記箇所などを始め、多くの技術／一般用語を含んでおり、また、一般用語のなかには審決特有の表現に組み込まれているものも多数存在する。こうした用語に関しては、コーパスの学習による改善が見込まれる。

最終学習後の内容伝達レベルの評価では、両エンジンとも評価対象 100 文のうち 97 文にレベル 4～5 という高評価が付された。つまり大半の機械翻訳文が「原文中の重要情報の 80%以上が正確に伝達されている」と評価されており、原文中の技術／一般用語も総じて高い精度で翻訳されていると見なせる。カテゴリ別エラーカウントでも「技術／一般誤訳」の最終的なカウント数は全体の 10%程度に抑えられており（汎用 NT:11 文、特許 NT:13 文。⇒4.5.5.2.）、この結論を裏付けている。

技術／一般用語の翻訳品質は、エンジンにあらかじめ施されている基礎学習と中国審決コーパスデータを介した学習、さらには大規模アップデート（⇒2.2.3.）による訳語選定ロジックの改良など、複数の要素が複合的に作用した結果である。各要素の比重がどの程度であったかを分別することは難しいが、少なくとも大規模アップデート後の「みんなの自動翻訳」に本事業で作成した中国審決コーパスデータ全件を学習させれば、技術／一般用語に関しても十分に高い翻訳精度が得られることが確認できた。

ただし、従来の機械翻訳で辞書中に存在しない用語が翻訳不可能であったのと同様、ニューラル機械翻訳においても、学習データ中に適切な文例が存在しない用語は正しく翻訳す

ることはできない。今回の調査分析により、「みんなの自動翻訳」に中国審決コーパスデータを学習させることで、現時点では、技術／一般用語、特許・審決用語ともおおむね網羅的にカバーできることがわかった。ただし、技術用語に関しては、日々新たな技術が発明され、それに伴い新しい用語が生み出されていく性質のものである。とりわけ特許関連文書である中国審決では、こうした新規の技術用語が早々に使用されることになる。このため、審査書類システムの実用にあたっては、最新の技術用語に対応できるよう、学習データを定期的にアップデートしていく必要がある。「みんなの自動翻訳」の定期的なアップデートが技術用語の訳質にさほど変化をもたらさなかったことを考えると、将来にわたり最新の技術用語へシステムを対応させていくためには、そのための学習データを定期的に補充することが望ましい。ただし、その際は、新規の技術用語が使用された中国特許由来の対訳コーパスを学習させる方が学習効果が高いと考えられる。

5.4. 技術用語・一般用語の不安定な変化について

第3.4.項や第4.5.4.項に示した実例の定点観測においては、技術用語や一般用語の中には、学習の都度あたかも複数の訳語候補からランダムに選定されているかのような予測不可能な変化を繰り返すものが散見された。さらには第3.4.項の実例3-3のように、同一文中で複数回用いられる技術用語（「破乳步骤」）に対して異なる訳語が採用されているケースも検出された。技術／一般用語におけるこうした傾向は、学習初期からコーパス由来の適訳語で常時安定した特許・審決用語とは大きく異なる。

この傾向の違いは、特許・審決用語はその大半が中国審決コーパスデータにより初めて学習されたものであり、学習データ中にコーパスで統一的に使用された単一の訳語候補しか存在しないのに対し、技術用語や一般用語は複数の訳語に訳し分ける必要があるものが多く、複数の訳語候補が学習データ（基礎学習データを含む）中に存在していることに起因すると考えられる。ニューラル機械翻訳では、学習データ中で一つの語（原語）が複数の訳語に訳し分けられている場合、それぞれの原文から得たその語の周辺情報に基づいて、各訳語にベクトルを付与する。そして、その語（原語）を含む文が入力されると、入力文におけるその語の周辺情報を学習データ中の各訳語のベクトルと照合して、最もフィットするベクトルを持つ訳語を選定している。このため、複数の訳語候補を持つ技術／一般用語では、同じ原語であっても入力文の内容によって異なる訳語が採用される場合がある。つまり同一用語の訳ゆれは、「みんなの自動翻訳」に限らず現在のニューラル機械翻訳における共通の課題といえる。本調査分析で多くの文の技術／一般用語の訳語が学習前後で変化を繰り返したのも、学習データが追加される度にベクトルが更新され、それによって複数の訳語候補間の優先順位が変化したためと考えられる。

「みんなの自動翻訳」においては、調査期間中に実施された「みんなの自動翻訳」の大規

模アップデート以降、学習前後の技術／一般用語の訳語変化はかなり抑制された。これは、エンジンのロジックが改良され、各語に付与されるベクトルが学習データの内容をよりの確に反映したものとなった結果、学習前後でのベクトルの安定性が増したためと推察される。とはいえ、大規模アップデート以降も一部の技術／一般用語の訳語変化は最終学習後まで生じており、これがニューラル機械翻訳方式を採用する「みんなの自動翻訳」の本質的な傾向であることには変わりがない。

こうした傾向による技術／一般用語の訳語変化が実用上の問題となるのは、学習前後での変化よりもむしろ同一学習回における、同じ技術用語の訳ゆれである。一件の中国審決は多数の文で構成されており、かつ、一案件内では複数の文の間で同じ技術用語が繰り返し使用されている。その技術用語が同一案件内で個々の文ごとに訳ゆれしていると、読み手を混乱させ、内容理解を阻害する。同一文中で「破乳步骤（解乳化ステップ）」の訳ゆれが発生していた事例 3-3 や第 4.5.5.10.項に示した文番号#1146 では、大規模アップデート以降は訳ゆれが解消していたが、ニューラル機械翻訳の訳語選定メカニズムを考えれば、これら二例をもって同一学習回での訳ゆれの懸念なしとはいえない。そこで本項において、最終学習後の一案件内の個々の文間での技術用語の訳ゆれの有無について実例を用いて検証する。

本調査分析に使用した評価用コーパスは、原則として案件単位の中国審決から選定されている（⇒2.4.1.）。このため、同一案件由来の各文に含まれる技術用語の訳語の一致・不一致を調査することが可能である。そこで、第 3.4.4.項で事例 3-4 として使用した文番号#1995 が属する中国審決 [復審：決定番号 77695] 由来の全ての文（文番号#1949～#2013 の 64 文）を対象に、出現頻度の高い技術用語 4 語、具体的には#1995 にも含まれていた「学员」「学习课件」⁵⁰「监控参数」「个性化」について、最終学習後の各機械翻訳文で採用された訳語を調査することとした。以下、その結果を示す。なお、調査対象の技術用語を含まない文は表から省略している。

表 5.4.1. 中国審決#77695 における技術用語の訳語状況（第 20 回後、汎用 NT）

原語	学员	学习课件	监控参数	个性化
正解訳	学生	学习コースウェア	監視パラメータ	個別化
#1950	—	—	—	個別化
#1954	—	—	—	個別化
#1955	—	運転学習教科書	—	—
#1956	訓練生	—	—	—

⁵⁰ #1995 に含まれる四語の残り二語「匹配的」「模拟练习」はいずれも出現が 2 文のみにつき掲載を省略した。なお、両語とも二文での訳語は一致していた。

#1957	学習者	—	—	—
#1958	訓練生	運転学習コースウェア	—	—
#1959	学習者	—	監視パラメータ	—
#1960	学習者	—	—	—
#1962	—	運転学習コースウェア	—	—
#1963	訓練生	運転学習コンテンツ	監視パラメータ	—
#1964	—	—	—	カスタマイズ
#1965	訓練生	—	監視パラメータ	—
#1966	—	—	—	カスタマイズ
#1967	訓練生	運転学習教材	—	—
#1968	学生	—	—	—
#1969	—	—	—	カスタマイズ
#1970	訓練生	運転学習教材	—	—
#1971	—	—	—	カスタマイズ
#1972	—	—	—	カスタマイズ
#1973	訓練生	—	—	—
#1974	訓練生	—	—	—
#1975	—	—	—	個別化
#1977	学生	—	—	—
#1978	—	—	—	個別化
#1979	訓練生	—	—	—
#1980	学習者	—	—	—
#1981	訓練生	運転学習パッケージ	—	—
#1982	訓練生	—	—	—
#1983	訓練生	—	—	—
#1984	訓練生	—	—	—
#1985	訓練生	—	監視パラメータ	—
#1986	訓練生	運転学習コース	—	—
#1987	—	運転学習教材	—	—
#1988	訓練生	運転学習コンテンツ	監視パラメータ	—
#1989	—	—	—	個別化
#1990	訓練生	—	監視パラメータ	—
#1991	学生	運転学習コース	—	個別化
#1992	訓練生	運転学習教材	—	—
#1993	訓練生	—	—	—

#1994	学生	運転学習コース	—	個別化
#1995	訓練生	運転学習教材	—	—
#2003	—	—	—	カスタマイズ
#2009	—	—	監視パラメータ	パーソナライズされた

表 5.4.2. 中国審決#77695 における技術用語の訳語状況（第 20 回後、特許 NT）

原語	学员	学習課題	监控参数	个性化
正解訳	学生	学習コースウェア	監視パラメータ	個別化
#1950	—	—	—	パーソナライズ
#1954	—	—	—	パーソナライズされた
#1955	—	運転学習教材	—	—
#1956	訓練生	—	—	—
#1957	訓練生	—	—	—
#1958	訓練生	運転学習教材	—	—
#1959	訓練生	—	監視パラメータ	—
#1960	訓練生	—	—	—
#1962	—	運転学習教科書	—	—
#1963	訓練生	運転学習教材	監視パラメータ	—
#1964	—	—	—	パーソナライズされた
#1965	訓練生	—	監視パラメータ	—
#1966	—	—	—	パーソナライズされた
#1967	訓練生	運転学習教材	—	—
#1968	訓練生	—	—	—
#1969	—	—	—	パーソナライズされた
#1970	訓練生	運転学習教材	—	—
#1971	—	—	—	パーソナライズされた
#1972	—	—	—	パーソナライズされた
#1973	訓練生	—	—	—
#1974	訓練生	—	—	—
#1975	—	—	—	パーソナライズされた
#1977	訓練生	—	—	—
#1978	—	—	—	パーソナライズされた
#1979	訓練生	—	—	—
#1980	訓練生	—	—	—
#1981	訓練生	運転学習教材	—	—

#1982	訓練生	—	—	—
#1983	訓練生	—	—	—
#1984	訓練生	—	—	—
#1985	訓練生	—	監視パラメータ	—
#1986	訓練生	運転学習教材	—	—
#1987	—	運転学習教材	—	—
#1988	訓練生	運転学習教材	監視パラメータ	—
#1989	—	—	—	パーソナライズされた
#1990	訓練生	—	監視パラメータ	—
#1991	訓練生	運転学習教材	—	パーソナライズされた
#1992	訓練生	運転学習教材	—	—
#1993	訓練生	—	—	—
#1994	訓練生	運転学習教材	—	カスタマイズされた
#1995	訓練生	運転学習教材	—	—
#2003	—	—	—	パーソナライズされた
#2009	—	—	監視パラメータ	個別化された

調査結果は上二表に示したとおりである。選定した4語のうち「监控参数」に関しては、両エンジンの全ての文で「監視パラメータ」という同じ訳語が採用されており、文間の訳ゆれは生じていない。しかしながら、残る3語、すなわち「学员」、「学习课件」及び「个性化」に関しては、同一学習回であっても、文によって複数の異なる訳語が採用されており、懸念していた同一案件内での訳ゆれが発生していた。

「学员」に関しては、「訓練生」という訳語が最有力候補となっており、特許NTでは全ての文(28文)でこの訳語が採用されている。ただし、汎用NTでは、「訓練生」が採用された文は20文にとどまり、残る8文では「学習者」「学生」という別の訳語が4文ずつで採用され、複数種複数回の訳ゆれが発生している。

「学习课件」は、汎用NTにおいてさらに激しい訳ゆれが生じた。全14文で採用された訳語は「運転学習教材」が5文、「運転学習コース」が3文、「運転学習コースウェア」と「運転学習コンテンツ」が各2文、「運転学習教科書」と「運転学習パッケージ」が各1文と、全6種類に上る。一方、特許NTでは「運転学習教材」が14文中13文を占め、訳ゆれは「運転学習教科書」が採用された1文のみであり、「学员」と同様、訳ゆれの程度は汎用NTよりもかなり軽微であった。

これら「学员」と「学习课件」の2語に関しては、訳ゆれは発生しているものの、最も多

くの文で採用された訳語は両エンジンで共通していた。しかしながら、「個性化」に関しては、汎用 NT では全 14 文中「個別化」が 7 文、「カスタマイズ」が 6 文と拮抗し、「パーソナライズ」が 1 文のみで採用されていたのに対し、特許 NT では「パーソナライズ」が 12 文と支配的で、「個別化」と「カスタマイズ」は各 1 文のみで採用されるという、最有力訳語候補のエンジン間での変化も生じている。

本調査結果からは、「みんなの自動翻訳」では、大規模アップデート以降は安定化が進んだとはいえ、技術用語の文間の訳ゆれは同一学習回であっても少なからず発生するものと結論せざるを得ない。

この訳ゆれは、学習データ中に存在する複数の訳語候補からその都度最善の訳語を採用するというニューラル機械翻訳の訳語決定のメカニズム⁵¹それ自体が内包する課題である。技術用語は技術内容に応じて複数の訳語に訳し分けられることが多く、特許・審決用語のように学習データ中の訳語を統一することは事実上不可能であるため、現状ではこの訳ゆれを完全に解消することは困難と考えられる。ただし、「みんなの自動翻訳」においては、大規模アップデート以降、訳語の安定性が明らかに向上し、大規模アップデート以前に見られていた学習前後の訳ゆれが大幅に解消された。今後もさらなる機械翻訳技術の進展により、文間の訳ゆれに関しても改善が進むものと期待される。

5.5. 「専利」の訳し分けに関する課題

「専利」は、中国における特許、実用新案、意匠の総称であり、対応する日本語が存在しない。このため中国審決コーパスデータ中では、案件の内容に応じて「専利」を「特許」と「実用新案」とに訳し分けた。だが、その結果、このコーパスを学習させた「みんなの自動翻訳」において、「専利」の「特許」と「実用新案」との間での訳し分け誤りが発生するようになった。

「みんなの自動翻訳」で「専利」が「特許」と「実用新案」とに訳し分けられるようになったのは、中国審決コーパスデータを介して「特許」と「実用新案」という複数の訳語が学習されたためである⁵²。学習データ中に複数の有力な訳語候補が存在することで、前述したニューラル機械翻訳の訳語選定メカニズムにより、技術／一般用語と同様の訳ゆれが発生したものと考えられる。

⁵¹ 詳しくは第 3.4.1.項を参照。

⁵² デフォルト状態の「みんなの自動翻訳」では「専利」は「特許」とのみ訳されており、「実用新案」と訳されるケースは存在しなかった。

本事業で翻訳対象とした案件は実用新案よりも特許のほうが圧倒的に多く⁵³、単純な出現頻度に基づけば「専利」の訳語は「特許」が断然優勢となる。しかしながら、ニューラル翻訳方式では、各文の訳語はそれぞれの文における周辺の語句の状況に基づいてその都度選定されるため、文によっては「専利」が「実用新案」と訳されることもある。その結果、的確に訳し分けられることも当然あるが、その一方で、「特許」と訳すべきところ「実用新案」と訳される、又は「実用新案」と訳すべきところ「特許」と訳されるという訳し分け誤りも発生している。

具体的な発生状況を調査したところ、評価用コーパス 1,994 文中、原文に「専利」を含み、かつ、正解訳文に「特許」を含み「実用新案」を含まない 556 文のうち、最終学習後の機械翻訳文に「実用新案」を含む文は汎用 NT で 144 文、特許 NT で 181 文存在しており、約 25～33%の文で「専利」の訳し分け誤りが発生していると考えられる。同じく原文に「専利」を含み、かつ、正解訳文に「実用新案」を含み「特許」を含まない 35 文のうち、最終学習後の機械翻訳文に「実用新案」を含まない文は汎用 NT で 8 文、特許 NT で 3 文と、こちらも約 9～23%の文で訳し分け誤りが生じている。

「専利」のように学習データ中に複数の有力な訳語候補が存在する語の訳し分けを正確に行うには、ニューラル機械翻訳技術のさらなる技術発展が必要であろう。それまでの間の暫定的な対処策としては、学習データから「専利」が「実用新案」と訳されている文を一時的に除外することで、「専利」の訳語を「特許」に固定することが可能と考えられる。本来「実用新案」と訳されるべき「専利」も一律「特許」と訳されてしまうことになるが、現状のランダム的な訳ゆれは解消できる。本事業で作成した中国審決コーパスデータは重複文を含めて約 300 万文だが、このうち「専利」を「実用新案」と訳している文は約 7 万文であり、本調査分析での 1 回分の学習データ量に満たない。学習後期の自動評価スコアや内容伝達レベルの状況から、仮にこれら 7 万文を除去したとしても翻訳品質への悪影響はごく小さいと考えられる。

⁵³ 本事業の翻訳対象案件 19,182 件中、特許が 16,981 件、実用は 2,201 件。

5.6. 長文における翻訳文の途絶について

「みんなの自動翻訳」における現状の課題としては、第 3.4.6.項の実例 3-6 で示したように、最終学習後の時点でも長文における機械翻訳文の途絶が発生していることも挙げられる。具体的な文字数など厳密な発生条件は特定不可能であったが、原文が極端な長文である場合、「みんなの自動翻訳」では翻訳文の途絶が生じる可能性が高い。

機械翻訳では、このように長文において翻訳文の途絶が多発する場合、機械翻訳システムの性能、具体的にはバッファの上限文字数を超過した結果である可能性が高い。この場合、ユーザが学習データ等を介して改善することは不可能である。

ただし、本調査期間中においても大規模アップデート以降、「みんなの自動翻訳」の長文処理能力は調査開始当初より明らかに改善された。途絶は依然として解消しないものの、各文とも出力される文字数が顕著に増加しており、アップデートを機にシステムのバッファが強化されたものと推察される。今後もハード・ソフト両面の技術発展に伴い、長文への対応もさらに改良されていくものと期待される。

5.7. 汎用 NT における日付エラーについて

「みんなの自動翻訳」における現状の課題としてはもう一点、第 4.5.5.8.項で報告した汎用 NT における「日付の日部分が異なる日付に変化するエラー」が挙げられる。

このエラーは、原文中の日付の日部分が翻訳文中で異なる日付（多くの場合は 1 日）に誤訳され、本来の日付は文末に数値として出力される、というエラーである。例えば文番号#1078 の最終学習後の機械翻訳文では、原文の「2017 年 03 月 06 日」という日付が「2017 年 3 月 1 日」となり、かつ翻訳文の文末に「06」という本来の日付が数値で出力されている。

本エラーは第 19 回学習後から発生したものであり、評価用コーパスでの状況を見る限り、汎用 NT のみで発生している可能性が高い。日付を含む文でも発生するものとしなものがあり、かつ、同じ文でも第 19 回後には発生せず第 20 回後のみ発生しているものもある。エラーが発生した場合、日付は「1 日」に置き換わることが多いが、文番号#1517 では「7 月 7 日」が「7 月 8 日」に置き換わっているなど、発生メカニズムは不詳である。

とはいえ、エラーの内容や、一方のエンジンのみで第 19 回後から突如発生している状況から、学習データすなわち中国審決コーパスデータの問題とは考えにくく、エンジンのロジック、又は前処理／後処理に起因したものである可能性が高い。

5.8. 汎用 NT と特許 NT の優劣

本調査では「みんなの自動翻訳」で利用可能な2つの翻訳エンジン、「汎用 NT」と「特許 NT」の双方を対象とした。二者の違いは主にエンジン構築のための基礎学習データの違いであり、具体的には、特許 NT が特許文献由来の対訳コーパスのみを使用しているのに対し、汎用 NT は汎用の対訳コーパスと、特許 NT と共通の特許文献コーパスの一部とを併用している（⇒2.2.1.）。無学習時はこの違いが顕著に表れており、例えば汎用 NT では未知語扱いされた「技術方案」が特許 NT では「技術的解決手段」と正しく訳出されるなど、一部の特許用語について特許 NT のみで適切な訳語が採用されていた。その一方で、特許文献由来のコーパスのみを重点的に学習したことによる副作用も若干数検出された（⇒4.5.5.7.）。

だが、中国審決コーパスデータの学習の結果、中国審決で使用されるほぼ全ての特許・審決用語がごく早期から安定してコーパスに準拠した適訳語で訳されるようになり、両エンジンの翻訳品質に特段の差異は見いだせなくなった。全体平均スコアでの比較となる BLEU や RIBES の自動評価及び内容伝達レベルの評価のいずれにおいてもコーパス学習後は両エンジンの差異は誤差レベルの僅かなものとなり⁵⁴、人手評価の一環として実施した「両エンジン間の機械翻訳文の相対評価」においても、学習を重ねるごとに個々の文単位でも「両者同等の品質」と評価される文が大勢を占めるようになった（第20回時は100文中79文）。

なお、技術／一般用語の翻訳品質は第5.3.項で述べたとおりエンジンの基礎学習に依存する部分が多い。汎用 NT と特許 NT では基礎学習に使用したデータが一部異なることから、技術／一般用語の翻訳品質に優劣が存在する可能性があった。しかしながら、上記のとおり中国審決コーパスの学習後は自動評価、人手評価とも両エンジンの翻訳品質は常に同等であった。本報告書に示した各実例においても、機械翻訳文に採用される有力な訳語候補は両エンジンほぼ共通しており、例えば特定の訳語が一方のエンジンのみで安定的に採用される等の、基礎学習データの違いを示すような事例は皆無であった。こうした状況から、技術／一般用語の翻訳品質に関しても、両エンジン間に顕著な優劣は存在しないといえる。

したがって、特許・審決用語、技術／一般用語とも、中国審決コーパスデータ全件を学習させる前提であれば、汎用 NT と特許 NT のどちらのエンジンを使用しても得られる品質は同等であると結論される。

ただし、前項で述べたとおり、現「みんなの自動翻訳」（GPMT-3.9_210930_nmt）では「日付の日部分が異なる日付に変化するエラー」が発生しており、このエラーは汎用 NT の

⁵⁴ コーパス学習初期には汎用 NT が特許 NT よりも若干優位であり、また無学習時においても内容伝達レベルは汎用 NT のほうが上位となったが、これは主に特許 NT が基礎学習に特許文献由来のコーパスのみを重点的に使用したことによる副作用（⇒4.5.5.7.）が主な原因と考えられる。

みで発生するものである可能性が高い点に留意が必要である。

5.9. 自動評価指標における第2回学習以降のスコアの停滞について

自動評価指標 BLEU、RIBES の全体平均スコアは、第1回学習後に大幅な向上を見せたが、第2回後以降は（第12回後を除き）微増と微減を繰り返すようになり、飽和とも映る様相を呈した。人手による内容伝達レベルの評価も同様であり、このことから、学習データは学習初期の数十万文対程度が適量と見なすこともできる。

ただし、一般に学習データ量の適量は機械翻訳エンジンにより異なり、また、同じ機械翻訳エンジンであってもバージョンによってその適量は変化する。本調査分析においても、第2回学習後以降の各回が横ばい傾向を見せるなか、「みんなの自動翻訳」の大規模アップデートが実施された直後の第12回後のみは顕著なスコア向上を示し、以降の各回でもそのスコア水準が維持された。人手評価や実例の定点観測から、ここでの品質改善は、学習データにおける出現頻度が訳語選定によりの確に反映されるようになったことで、突発的な不適訳語の採用が減少したことが主な理由であることがわかった。ニューラル機械翻訳における訳語選定のメカニズムは、前述のとおり、学習データ中に存在する訳語に周辺情報から多次元ベクトルを付与し、入力文の状況に応じて訳し分けるというものである。このベクトルは学習データ量が多いほど精密なものになるため、「みんなの自動翻訳」の大規模アップデートも、第12回当時の学習データ量でなかった場合、同等の改善効果が得られたとは限らない。

今後もソフト・ハード両面の技術発展に伴い、学習できるデータの規模は拡大していくと考えられる。将来の中国審決の機械翻訳の実用化に向けては、特定の機械翻訳システムでの現状での飽和点にとらわれず、高精度の対訳コーパスを大量に作成して備えておくべきである。本事業では、現時点での機械翻訳システムに用いる学習データの整備という観点では十分に目的を達成したものの、機械翻訳技術の近年の発展は急速であり、その点は今後も注意が必要である。

また、本事業で作成した中国審決コーパスデータは、あくまで現在の中国審決の記載内容に対応するためのものである。中国の復審請求や無効宣告請求の増加が予測される中、審決の記載内容は、今後の法改正、法解釈を新たにする画期的な判決の登場、新たな技術分野への対応により、時代とともに変化していくことが予想される。特に2021年6月に施行された改正中国専利法には、懲罰的賠償の導入、法廷賠償金額の引き上げ、医薬品特許の存続期間延長制度の新設などが含まれ、当法改正により新たなタイプの審決文が増えることが予測される。このため、現在と同等の機械翻訳品質を長期にわたり維持するためには、今後も適切なタイミングで最新の中国審決のコーパスデータを適量作成し、これを追加学習させ

ていくことが望ましい。

5.10. 本事業で達成された中国審決の機械翻訳品質について

本調査では、本事業で作成した中国審決コーパスデータを機械翻訳システム「みんなの自動翻訳」に学習させることにより、中国審決の機械翻訳品質がどの程度改善されるかを検証した。その結果、全体としては「重要情報の 80%以上が正確に伝達される」レベルの機械翻訳精度が達成されることが確認された。

中国審決は、例えば審決の結論や適用条文など、定型的な文パターンで述べられる箇所が多い。この特徴は、学習データの各文例における語同士の関連性に基づいて訳語を決定するニューラル翻訳方式に適しており、今回の調査分析結果からも、少なくともこれら定型的な文が用いられる箇所に関しては、中国審決コーパスを学習データとすることで、正確な翻訳文を提供することが可能であると結論できる。

一方、全件学習後も文によっては誤訳等の不備が残存しているものもあり、また、同一案件内における技術用語の訳ゆれや長大な文における翻訳の途絶など、コーパスの学習のみでは解決が困難と思われる課題もいくつか検出された。しかし、現在の機械翻訳において 100%の精度を期待することは非現実的であり、機械翻訳サービスを利用するユーザもそのことは十分に承知している。

現在、中国審決情報は中国語でのみ提供されており、これに日本語で簡便にアクセスする手段は存在しない。本事業で作成した中国審決コーパスデータは、学習データとして「みんなの自動翻訳」に供することにより、中国審決を「重要情報の 80%以上が正確に伝達される」品質で日本語に機械翻訳することを可能とした。本コーパスを活用し、これを学習した機械翻訳エンジンで中国審決の和訳文を提供することで、中国審決への日本語によるアクセス性は大きく改善されるであろう。中国審決の内容を十分な精度で簡便かつ迅速に把握できるようになり、我が国ユーザの中国における権利取得の業務の効率化に大いに貢献するものと考えられる。