ERUDAITE · CATER V2 TRANSLATION DIAGNOSTIC

CATER v2 Report (Advanced) — 2026-05-13

MODEL	TEMPERATURE	ORCHESTRATOR	EVALUATED
openai/gpt-5.4	0	cater-v2.orchestrator.2026-05-10b	2026/5/13 11:59:31

OVERALL SCORE

96.9 / 100 A — Production-ready

モデル	openai/gpt-5.4
Temperature	0
評価基準	621 TW
BriefStatus	inferred
Confidence	low
評価日時	2026/5/13 11:59:31

全体として、候補訳はSource Modelが示す学術的・技術的内容を高水準で再現しており、Target Modelが要求する専門的でフォーマルな日本語にも概ね適合している。もっとも重要な問題はFC上の1件で、末尾文において「公開されていることが改良と検証を促す」という原文の命題関係が「推奨する」へと変質し、Obligation Matrixの命題不変条件を損なっている点である。加えて、評価軸名の訳語に軽微なTC上の揺れがあり、専門分野で期待される用語体系の安定性がやや弱い。さらに、「価値のあるツールとして登場した」はCSAの観点で学術日本語としての自然さを欠き、CATERの有用性を確信させるという効果不変条件の実現を弱めている。

AXIS SCORES

GP 文法的精度 — 100.0

重み 13.0%

エラー	EUC	REUC
0	0	0.0

EditEffortER	RiskAdjER	RawScore
0.00	0.00	100.0

Sens. 1.00

SI 意味整合性 — 100.0

重み 21.7%

エラー	EUC	REUC
0	0	0.0

EditEffortER	RiskAdjER	RawScore
0.00	0.00	100.0

Sens. 1.60

FC 事実整合性 — 88.4

重み 21.7%

エラー	EUC	REUC
1	5	30.0

EditEffortER	RiskAdjER	RawScore
0.81	4.83	88.4

Sens. 2.40

TC 用語整合性 — 97.5

重み 21.7%

エラー	EUC	REUC
2	4	8.0

EditEffortER	RiskAdjER	RawScore
0.64	1.29	97.5

Sens. 1.90

DC 談話一貫性 **100.0**
重み 13.0%

エラー	EUC	REUC
0	0	0.0
EditEffortER	RiskAdjER	RawScore
0.00	0.00	100.0
Sens.		
1.50		

CSA 伝達・文体適切性 **99.0**
重み 8.7%

エラー	EUC	REUC
1	4	4.0
EditEffortER	RiskAdjER	RawScore
0.64	0.64	99.0
Sens.		
1.50		

Revision Priorities

1. [FC] 末尾文の命題関係を原文どおりに戻し、「公開されていること」がコミュニティ主導の改良と実証的検証を促進するという構図を正確に再現してください。
2. [TC] 評価軸名の訳語を専門分野でより安定した日本語に統一し、とくに「semantic fidelity」「contextual coherence」に対応する用語を学術的慣用に合わせて整備してください。
3. [CSA] 「価値のあるツールとして登場した」のような直訳調の表現を、日本語の学術的紹介文として自然で説得力のある言い回しに置き換え、読者に対する提示効果を高めてください。

UNCERTAINTY NOTES

- 「CATER」の日本語名称（包括的AI支援翻訳編集率）は、定訳がないため原文の直訳として妥当性を判断する。
- 「Edit effort」の訳語として「編集作業量」が最適か、あるいは「編集努力」などの表現が好まれるかは文脈に依存する。
- [S1 全文] 末尾文の「公開されており、…推奨する」に関するFC・DC・CSAの3件は、同一の基底問題を異なる軸から捉えた重複指摘と判断し、命題関係の変質を最小概念修正として要するFCに統合した。EUCは統合元のうち最大の5を維持した。
- [S1 全文] 「価値のあるツールとして登場した」に関するCSAとGPの2件は、主たる問題が学術文体への不適合にあるためCSAへ統合した。GPは表層的な不自然さとして二次的影響に回した。
- [S1 全文] 「openly available」→「公開されており」に関するTC指摘は、与えられたReferenceから必須の定訳違反までは立証しにくく、また同箇所の主要問題は別途FCで処理済みであるため、独立エラーとしては採用しなかった。条件付きの改善提案にとどまるため、統合段階で削除した。
- 「semantic fidelity」の最適訳としては『意味忠実性』が有力であるが、分野や文脈によっては近接する別訳も成立しうるため、最終的な用語選択は文書全体の術語方針との整合確認が望まれる。
- 「contextual coherence」については『談話的一貫性』への統一が妥当と判断したが、当該文書で文脈レベルと談話レベルをどこまで厳密に区別しているかは、本文全体がないため確定できない。

EVALUATION CONFIDENCE: LOW

- Target Model was inferred from source-side genre or situational cues; inferred assumptions are provisional evaluation assumptions, not user intent.
- Multiple unresolved interpretation or reference uncertainties were recorded across the pipeline.

Sampling Metadata

モード	サンプル数	評価対象 (原文)	評価対象 (訳文)
自動サンプリング	1	1,547 字	669 字
原文全文	サンプル比率	自動設定	評価基準
1,547 字	100.0%	5 × 1500字	621 TW

ERRORS (4)

s1-tc-001 TC · Terminological Consistency 軽微 Strong Default functional EUC 2 conf: medium MQM: terminology +CSA

評価軸名として用いられた「意味的忠実度」は、専門分野で一般的な日本語訳としてはやや不安定であり、術語の統一性を欠いている。内容理解を大きく損なうほどではないが、学術的な用語運用として精度が不足する。

Referenceでは、評価軸名は専門分野で一般的な日本語訳で表現することが求められている。該当箇所は命題内容そのものを改変するものではなく、主として用語選択の問題であるため、最小概念修正の観点からTCが主軸となる。Target Modelでも学術的に適切な用語使用が要求されており、その要件に対する軽微な不適合として扱うのが妥当である。

SOURCE	CANDIDATE
semantic fidelity	意味的忠実度

修正案
意味忠実性

s1-tc-002 TC · Terminological Consistency 軽微 Strong Default functional EUC 2 conf: medium MQM: terminology +DC

「contextual coherence」に対する訳語「文脈的一貫性」は、当該文脈で求められる専門用語体系との整合性がやや弱い。評価観点の名称としては、より安定した術語選択が望まれる。

Referencesでは「discourse-level」に「談話レベル」が対応付けられており、関連する評価概念群を日本語で提示する際には「談話」系の術語でそろえる方が用語体系上整合的である。本件は談話構造の破綻そのものではなく、評価ラベルの訳語選択に関する問題であるため、DCではなくTCを主軸とする。

SOURCE	CANDIDATE
contextual coherence	文脈的一貫性

修正案
談話的一貫性

s1-fc-001 FC · Factual Consistency 重大 Strong Default propositional EUC 5 conf: high MQM: mistranslation +CSA, DC

末尾文で、原文の「公開されていることが改良と検証を促す」という関係が、「改良と検証を推奨する」という別の作用に置き換えられている。公開状況から生じる誘発・促進の記述が変質しており、命題レベルの内容がずれている。

Source ModelおよびObligation Matrixの機能・命題要件では、フレームワークとプロンプト例の公開がコミュニティ主導の改良と実証的検証を後押しするという構図が保持される必要がある。候補訳は「encouraging」を「推奨する」としており、公開されている事実の帰結を記述する文から、何かが他者に勧奨する文へと命題関係を変更している。これは単なる文体や結束の問題ではなく、保護される命題内容の損傷であるため、決定木に従いFCを主軸とする。

SOURCE

```
The framework and example prompts are openly available, encouraging community-driven refinement and further empirical validation.
```

CANDIDATE

本フレームワークとプロンプトの例は公開されており、コミュニティ主導の改良とさらなる実証的検証を推奨する。

修正案

本フレームワークとプロンプトの例は公開されており、コミュニティ主導の改良とさらなる実証的検証を促している。

s1-csa-001 CSA · Communicative & Stylistic Appropriateness 軽微 Conditional effect EUC 4 conf: high MQM: register +GP

「価値のあるツールとして登場した」は、日本語の学術的紹介文として不自然で、原文の評価的提示を硬直した直訳調にしている。内容は概ね保持されているが、媒体と読者に対する文体適合性が弱い。

この箇所の主要な問題は、命題内容の誤りというより、Target Modelが求める学術的・専門的な日本語としての自然さと提示効果の不足にある。「emerges as a valuable tool」を「価値のあるツールとして登場した」とすると、日本語の技術文書としてはコロケーションとレジスターの両面で不自然さが残る。修正の中心は文体・表現選択であり、決定木上はCSAが主軸となる。GP由来の指摘は下流影響として統合するのが妥当である。

SOURCE

```
By uniting the conceptual rigor of frameworks like MQM and DQF with the scalability and flexibility of LLM-based evaluation, CATER emerges as a valuable tool for researchers, developers, and professional translators worldwide.
```

CANDIDATE

CATERは世界中の研究者、開発者、プロの翻訳者にとって価値のあるツールとして登場した。

修正案

CATERは世界中の研究者、開発者、プロの翻訳者にとって有用なツールとなる。

⚠ Reconstructed Context

SOURCE MODEL

Genre	学術論文の要旨（アブストラクト）または導入部。
Field	自然言語処理、特に機械翻訳の評価手法。
Tenor	専門的かつフォーマル。
Mode	公表を前提とした書き言葉。
Audience	機械翻訳の研究者、開発者、および翻訳実務者。
Function	新しい評価フレームワーク「CATER」の概念、利点、および有用性を説明し、普及を促す。
Textual Affect	厳密で信頼性が高く、革新性を強調するトーン。
Discourse Structure	定義、特徴、実装上の利点、対応可能な課題、既存手法との統合、公開情報の順で構成される。
Epistemic Status	非常に高い。特定の技術的提案を詳細に記述している。

TARGET MODEL · BRIEF: INFERRED

Audience	日本のNLP研究コミュニティ、AI開発者、LSP（翻訳会社）の品質管理担当者。
Medium	技術論文、学会発表資料、または専門的な製品紹介。
Function	原文の技術的詳細とニュアンスを正確に日本語で再現する。
Quality Regime	厳格な正確性と、学術的に適切な用語の使用が求められる。
Acceptable Loss	ほぼゼロ。技術的な定義や利点の記述において情報の欠落は許されない。
Required Effect	読者がCATERの革新性と実用性を正しく理解し、既存手法との違いを把握できること。
Permitted Reconstruction	日本語の学術論文として自然な表現（「本稿では」や「〜である」調）への調整。

SURFACE INVARIANTS

- CATER
- AI-assisted Translation Edit Ratio
- MT
- LLM
- MQM
- DQF

REFERENTIAL INVARIANTS

- CATER (Comprehensive AI-assisted Translation Edit Ratio)
- Large Language Models (LLMs)
- Machine Translation (MT)
- MQM (Multidimensional Quality Metrics)
- DQF (Dynamic Quality Framework)

PROPOSITIONAL INVARIANTS

- CATERはプロンプト駆動型のフレームワークである。
- 参照訳に依存しない（reference-independent）評価が可能である。
- LLMを活用してエラー特定、編集量定量化、スコアリングを行う。
- 事前計算された参照やドメイン固有リソースを必要としない。
- 重み付けやプロンプト変更により多様なニーズに適応可能である。
- 微妙な脱落、ハルシネーション、談話レベルのずれを捕捉できる。
- MQM/DQFの厳密さとLLMの柔軟性を統合している。

FUNCTIONAL INVARIANTS

- 学術論文の要旨として、新しいフレームワークの定義、利点、公開状況を明確に提示する。

• 専門用語を正確に使用し、技術的な信頼性を維持する。

EFFECT INVARIANTS

• 読者（研究者や開発者）に対して、CATERが革新的で実用的なツールであることを確信させる。
• 既存の評価指標（MQM, DQF）との関連性を示し、その進化形であることを理解させる。

AUTHORITATIVE GROUNDING

• MQM: Multidimensional Quality Metrics
• DQF: Dynamic Quality Framework
• Hallucination: ハルシネーション（または幻覚）
• Discourse-level: 談話レベル

NORMATIVE CONVENTIONS

• 学術論文の導入部として「本稿では（This paper introduces...）」などの定型表現を使用する。
• 「Linguistic accuracy」や「Semantic fidelity」などの評価軸を、専門分野で一般的な日本語訳で表現する。

APPENDIX・用語と理論

本 Appendix は、レポートを単体で読めるようにするための公開可能な参考情報です。CATER v2 の 6 軸定義・主要用語・スコア式・評価モード・方法論上の前提・ Translation Strategy Design Framework（TSDF）のハイレベル概念を収録しています。

Appendix S スコアの読み方（詳細版）

CATER v2 は、翻訳が **翻訳目的（Translation Brief）** にどれだけ適合しているかを **100 点満点**で算出します。

スコア帯と推奨アクション

スコア	品質判定	推奨アクション
90〜100	公開可能水準	最終校閲のみで利用可能
80〜89	業務利用可能	軽微な調整で実務に使えます
70〜79	部分修正が必要	該当箇所を特定して修正してください
60〜69	全面見直しが必要	致命的ではないものの、全編にわたる修正が必要です
50〜59	大規模な再作業が必要	翻訳目的への適合性に問題があります
40〜49	致命的な問題があります	用語規則違反、機能不全、文書構造の崩壊などの重大問題が含まれます
40 未満	使用できません	ハルシネーション、重大な事実誤認、致命的な誤訳が含まれます

軸別スコアの重要性

CATER v2 は次の 6 軸で翻訳を診断します。総合スコアが高くても、いずれかの軸で重大な問題がある場合があります。

- **GP（文法）**: 文法・表記の正確性
- **SI（意味）**: 意味の保存
- **FC（事実）**: 事実関係・固有名詞の正確性
- **TC（用語）**: 用語規則・表記ルールの遵守
- **DC（文脈）**: 文書全体の論理構成と一貫性
- **CSA（文体）**: 翻訳目的に対する文体・トーンの適切性

各軸は異なる速度で劣化します。たとえば文法（GP）は段階的に劣化しますが、事実関係（FC）は少数の誤りでも一気にスコアが落ちます。これは事実誤認が翻訳全体を使用不能にしうるためです。

「Cap」が示すもの

スコアレポートに **Cap トリガー** の表示がある場合、特定の重大問題が検出されたことを意味します。代表的なケース:

- 法務・医療・金融・安全関連の事実誤認 → 総合スコア 40 以下に制限
- 絶対的な機能要件の不充足 → 総合スコア 30 以下に制限
- 拘束的用語規則の反復違反 → 用語軸（TC）スコア 50 以下に制限

- 文書構造の崩壊 → 文脈軸（DC）スコア 50 以下に制限
- 翻訳目的の中核未達 → 文体軸（CSA）スコア 50 以下に制限

Cap がトリガーされた場合、スコアの数値そのものよりも **何が引き金になったか** を確認してください。

Brief の有無による信頼度

CATER v2 は **翻訳目的（Translation Brief）** が明示されている場合、最も高い精度で評価できます。Brief が未指定の場合、CATER v2 が自動推定しますが、推定された Brief に基づくスコアであることがレポートに明記されます。重要な業務利用では、Brief を明示することを推奨します。

Appendix A 6 軸（Six Axes）の定義

GP 文法的精度 Grammatical Precision

訳文が目標言語の表層として整っているか。

GP は、訳文が目標言語の文法・正書法・形態・句読法・表層的な自然さの面で整形されているかを評価します。意味の正しさそのものではなく、生成された目標言語表現が言語的に許容できるかを問います。

主な検出例

- 活用・呼応・格・時制などの文法ミス
- 誤字・脱字・スペルミス・句読点の誤り
- 形態論的な誤り（語形・送り仮名など）
- 明らかに不自然な語順や文構造の崩れ
- 目標言語としての流暢さの破綻

主軸となる場合

- 局所的な文法修正だけで問題が解消する場合は GP を主軸とします。

他軸との境界

- 文法は正しいが意味が誤っている場合は SI（または FC）。
- 文法は正しいが文体・レジスターが不適切な場合は CSA。
- 文法エラーが談話レベルの破綻を引き起こす場合、局所修正で解消するなら GP、文書全体の構造修復が必要なら DC。

MQM / DQF 対応

MQM の Fluency / Grammar / Spelling / Punctuation 系、DQF の流暢性に対応します。

SI 意味整合性 Semantic Integrity

原文の意味・含意・発話行為が訳文で保たれているか。

SI は、原文側で識別された意味（語彙的意味・概念関係・含意・発話行為・語用論的関係・意味の深さ）が訳文に保たれているかを評価します。辞書的意味だけでなく、要請・警告・約束・皮肉・含意・強調・対比など、原文の伝達行為に関わる意味も対象です。

主な検出例

- 語彙意味の誤訳・多義語の選択ミス
- 登場人物・参与者間の関係の取り違え
- 発話行為の誤り（依頼を提案にしてしまう等）
- 必要な含意の欠落・誤った保存
- 否定・条件・対比・強調の解釈ミス（保護される事実関係を破壊しない範囲で）
- 意図された意味のフラット化

主軸となる場合

- 誤訳の修正は意味の取り直しで解消し、保護される事実・参照・量・条件・義務的コミットメントを傷つけない場合。
- 原文側の意味・語用論的力を取り違えている場合は SI。

他軸との境界

- 誤りが保護される事実・参照・量・条件・義務的コミットメントを破壊する場合は FC。
- 誤りがバインディング用語集・公式名称に違反する場合は通常 TC。
- 問題が文単位ではなく文書レベルの連続性なら DC。
- 意味は把握されているが訳文として目標読者に効かないなら CSA。

MQM / DQF 対応

MQM の Accuracy / Mistranslation、DQF の Adequacy（妥当性）に対応します。

FC 事実整合性 Factual Consistency

保護されるべき事実・量・条件・義務関係が壊れていないか。

FC は、原文・Translation Brief・Reference 制約に由来する保護される、あるいはリスクを伴う事実的・参照的・数量的・条件的・義務的コミットメントを訳文が保っているかを評価します。LLM 時代の翻訳で問題化しやすい「ハルシネーション」「省略」「事実ドリフト」を一級の検出対象としてモデル化した軸です。

主な検出例

- 原文にない人物・属性・原因・条件・日付・数値などを足してしまう（ハルシネーション）
- 数値しきい値・安全条件・必須情報の欠落（省略）
- 「may」を「must」に強化する／その逆をする（モダリティのドリフト）
- 条件文を無条件に変える、責任主体を入れ替える
- 因果関係を時間関係にすり替える
- 法的・医療的・財務的に保護される情報の改変

主軸となる場合

- 改変された内容が「保護されるコミットメント閾値」（量・日付・通貨・条件・否定・因果・義務・許可・禁止・安全・法的効力など）を越えるとき。

他軸との境界

- 追加された情報が事実関係を壊さない場合は CSA / DC で扱う場合があります。
- 用語の不一致が事実参照を変えないなら TC。
- 意味の歪みが保護されたコミットメントを破壊しないなら SI。
- Translation Brief が要約を許可している場合の意図的な省略は FC のエラーとは扱いません。

MQM / DQF 対応

MQM の Accuracy / Omission / Addition / Untranslated に対応し、特にハルシネーション・事実ドリフトを一級扱いにします。

TC 用語整合性 Terminological Consistency

用語集・公式名称・ブランド規約・スタイル規則の遵守。

TC は、用語・固有名詞・領域固有表現・公式表記・クライアント指定の用語、および繰り返し現れる語彙選択が、適用される Reference 制約に従って一貫しているかを評価します。狭義の用語のみならず、UI ラベル・大文字化規則・転写規則など、Reference によって縛られた表層制約全般を含みます。

主な検出例

- ユーザー用語集・スタイルガイドへの違反
- 同一概念を文書内で異なる訳語で扱ってしまう
- 公式名称があるのに非公式表記を使う
- ブランド・製品・法人名の不統一
- 領域用語と一般語の取り違え
- 必須の大文字化・転写・正書法ルールの違反

主軸となる場合

- Reference（ユーザー用語集・公式名称・ブランド規約・スタイルガイド・既訳メモリ等）が有効・適用可能であり、より優先度の高い義務にも上書きされていない場合。

他軸との境界

- 用語の不一致が事実参照そのものを変えるなら FC を主軸、TC は副次タグ。
- Reference に縛られていない単なる意味の問題なら SI。
- 用語自体は適切だが対象読者に対して文体的に不適切な場合は CSA。

MQM / DQF 対応

MQM の Terminology / Locale Convention / Named Entity、DQF の用語管理に対応します。

DC 談話一貫性 Discourse Coherence

文書レベルの結束・首尾一貫性・情報構造を保っているか。

DC は、訳文が文書レベルの結束性（cohesion）・首尾一貫性（coherence）・情報構造・参照連続性・主題進行・談話論理を保っているかを評価します。文単位では文法的・意味的に正しくても、文書として読めない訳文を検出するための軸です。CATER v2 で CSA から独立させた診断次元です。

主な検出例

- 代名詞・照応関係の連鎖が崩れる
- 話題の進行が一貫しない
- 段落内の論理がつながらない
- 接続関係（対比・譲歩・原因・結果など）の誤りや欠落
- Given/New 情報構造が崩れた語順
- 視点・話者の追跡が乱れる
- 文単位は正確だが、まとまった文書として読みにくい

主軸となる場合

- 文単位の意味・事実は概ね正しいが、結束・首尾一貫性・照応・主題進行・段落論理が崩れているとき。

他軸との境界

- 1 文の意味そのものが誤っているなら SI または FC。
- 整合性は保たれているが対象読者に合っていないなら CSA。
- 文法エラーが原因で首尾一貫性が崩れている場合、局所修正で解消するなら GP。

MQM / DQF 対応

MQM の Coherence / Cohesion / Discourse-level Accuracy、DQF の文書レベル品質基準に対応します。

CSA 伝達・文体適切性 Communicative & Stylistic Appropriateness

対象読者・媒体・目的に対して伝達設計として機能しているか。

CSA は、訳文が「目標側の伝達イベント」として適切に機能しているかを評価します。Translation Brief と Target Model のもとで、意図した目標側の効果・スタイル・レジスター・トーン・文化的適合・読者適合が実現されているかを問います。原文の表層スタイルを機械的に再現することではなく、目標読者に対して同等または指示された効果を生むかどうかを見ます。

主な検出例

- 法務・投資家向け文書での過度にカジュアルなトーン
- 扇動・説得を要する場面での硬すぎる文体
- キャラクターの声・著者の立場の喪失
- 修辞的な力（説得・教示・感情・ユーモア・厳粛さ・緊迫感）の欠落
- ローカリゼーションの過剰／不足（過度な同化・過度な異化）
- 対象読者の知識量に合わない説明密度

主軸となる場合

- 整合性も意味も維持されているが、対象読者・媒体・機能・品質体制に対して効かない訳文になっているとき。

他軸との境界

- 訳文がそもそも結束していないなら DC。
- 意味そのものが誤っているなら SI または FC。
- 特定のスタイルガイド規則の違反が原因なら TC が主、CSA は副次タグ。

MQM / DQF 対応

MQM の Style / Register / Locale Convention / Audience Appropriateness、DQF の対象読者・目的別の品質基準に対応します。

Appendix B Severity と Obligation Strength

SEVERITY（誤りの深刻度）

Minor（軽微） ×1

目立たないが許容しがたい質的低下。読み手の理解や品質印象に影響するが、機能や事実は損なわれない。

Major（重大） ×3

OBLIGATION STRENGTH（保護強度）

Optional（任意） ×0.5

違反しても結果に大きな影響を及ぼさない、preference に近い制約。

Conditional（条件付） ×1

条件・領域・読者によっては守るべき制約。Translation Brief や領域慣習に依存する。

読み手の理解や品質に明確に影響する誤り。機能・効果が部分的に損なわれる、または訂正なしには利用に耐えない。

`Critical（致命的）` ×8

事実・安全・法的妥当性・運用妥当性などを破壊する誤り。そのまま公開すれば実害が想定される水準。

`Strong Default（強い既定）` ×2

明示的に上書きされない限り守るべき制約。専門領域・公式表記の多くがここに該当。

`Absolute（絶対）` ×5

Translation Brief・法令・契約・安全要件などにより、ほぼ例外なく遵守が要求される制約。

`Appendix C` スコア式（Edit Ratio パイプライン）

EUC Edit Units to Correct（編集単位）

個々のエラーを訂正するのに必要な最小編集量の見積り。語数や文字数ではなく「修正の手間」をスケールフリーに表現します。

REUC Risk-adjusted EUC

EUC に Severity 倍率と Obligation 倍率を掛け合わせ、リスクに比例した修正コストとして表したもの。CATER v2 のスコアリングの中核量です。

```
REUC_error = EUC × Severity × ObligationStrength
```

EvaluationBase 評価基準量

Edit Ratio の分母。Target Words（既定）または Source Words を用い、CJK は文字単位、ラテン文字系は空白区切りトークン単位で算出します。

EditEffortER 編集労力比

リスクを加味しない、純粋な編集量ベースの軸別 Edit Ratio。複数評価間の編集量比較に使えます。

```
EditEffortER_axis = ΣEUC_axis / EvaluationBase × 100
```

RiskAdjustedER リスク調整済 Edit Ratio

REUC を用いた、リスクを織り込んだ軸別 Edit Ratio。CATER v2 の最終スコアの直接の入力となります。

```
RiskAdjustedER_axis = ΣREUC_axis / EvaluationBase × 100
```

Sensitivity 軸感度

各軸が RiskAdjustedER の増加に対してどれだけ急峻にスコアを下げるかを決める係数。FC が最も急峻（2.4）、GP が最も緩やか（1.0）です。

RawScore 未補正スコア

AxisCap や GlobalGate を適用する前の軸別スコア。100 から RiskAdjustedER × Sensitivity を引いた値で、0 と 100 にクリップされます。

```
RawScore_axis = max(0, min(100, 100 - RiskAdjustedER × Sensitivity))
```

AxisCap 軸キャップ

致命的な FC・TC・DC・CSA エラーが検出された場合に、その軸の最終スコアに課される上限。RawScore がいくら高くても上限以下に抑えられます（§11.5）。

GlobalGate 全体ゲート

高ステークス領域での致命的 FC、絶対的 Functional Invariant の違反、Brief の目標機能を逸した致命的 CSA など、全体スコアそのものに上限を課す条件（§11.5）。

FinalScore 最終スコア

AxisCap 適用後の軸別スコアと、軸の重みから合算したスコアに、必要に応じて GlobalGate を適用した値（0–100）。

```
OverallScore = min(GlobalGate, Σ FinalScore_axis × weight_axis)
```

AxisVariance σ²(RiskAdjustedER) サンプル間分散

Advanced モードで複数サンプルを評価したときの、サンプル毎 RiskAdjustedER の分散。値が大きいほど文書内の品質ムラを示唆します。

スコア計算は CATER v2 §11 に準拠して決定論的に行われ、LLM には委譲されません。致命的な誤りに対する AxisCap・GlobalGate は §11.5 に従い、最終スコアの上限を直接抑えます。

Appendix D 評価コンテキストの用語

Source Model 原文モデル

原文を「ジャンル・領域・テナー・モード・読者・機能・テクスト的情動・談話構造・意味内容・認識的地位」といった次元に分解した記述。CATER v2 はこれを根拠に評価判断を行います。

Target Model 目標モデル

訳文が満たすべき条件（読者・媒体・機能・品質体制・許容される損失・必要とされる効果・許される再構成）を記述したもの。多くは Translation Brief に由来します。

Translation Brief 翻訳ブリーフ

翻訳依頼に伴う指示・条件のまとまり。読者・媒体・目的・品質体制・許容される損失・必要とされる効果などを含み、CATER v2 の評価アンカーとなります。明示・部分的・推定・不在の 4 段階で扱います。

Reference Layer 参照層

用語集や公式名称、領域慣習、既訳メモリなど、訳文が遵守または参照すべき外部資料。指示的（Directive）／権威的（Authoritative）／規範的（Normative）／経験的（Empirical）の 4 階層で扱います。

Obligation Matrix 義務マトリクス

訳文に保たれるべき不変条件（Surface / Referential / Propositional / Functional / Effect）と、その保護強度（Optional〜Absolute）を整理した枠組み。誤りの主軸選択と Obligation 倍率の根拠になります。

Brief Status ブリーフ状態

Translation Brief が explicit（明示）／ partial（部分）／ inferred（推定）／ absent（不在）のいずれであるか。inferred / absent では特に CSA の判断が低信頼として扱われます。

Appendix E 評価モード

簡易版（Basic / Quick）

原文と訳文を 1 回の LLM 呼び出しで評価し、軸別スコアと短評を返す軽量モード。スクリーニングや初期トリアージ向け。

簡易版のスコアは LLM が直接判定した推定値であり、§11 の REUC / RiskAdjustedER パイプラインを経たものではありません。学術検証や論文発表にはアドバンスド版を使用してください。

アドバンスド版（Advanced / Auto）

入力規模に応じて自動でサンプル数（N=3 / 5 / 10）を決定し、Stage 1〜7 のフルパイプラインを通します。Source/Target Model 構築、Reference / Obligation Matrix 抽出、6 軸並列評価、サンプル統合、戦略診断までを行います。

サンプリングは段落／文末に自動スナップされ、複数サンプル間の集計は REUC・EvaluationBase の合算を通じて intensive-quantity（強度量）として保持されます。

Appendix F Translation Strategy Design Framework（TSDF）

TSDF は、翻訳を「原文側の伝達イベント（Source Model）」から「目標側の伝達イベント（Target Model）」への、制約付きの変換として捉える事前設計フレームワークです。CATER v2 はこの事前設計に対する事後診断レイヤーとして設計されており、両者は同じ用語体系を共有します。

Source Model — 原文の伝達イベント記述

原文をジャンル・領域・テナー・モード・読者・機能・テクスト的情動・談話構造・意味内容・認識的地位といった次元で記述します。

Target Model — 目標の伝達イベント設計

翻訳が達成すべき読者・媒体・機能・品質体制・許容される損失・必要とされる効果を、Translation Brief に基づいて設計します。

Reference / Obligation — 制約と保護対象

用語集・公式名称・領域慣習・既訳メモリといった参照資料と、Surface / Referential / Propositional / Functional / Effect の不変条件群が、変換に課される制約として機能します。

Strategy — 解釈と表現の方針決定

Source Model の解釈と Target Model に向けた表現方針を、上記の制約とリスク前提のもとで設計します。

CATER v2 は、TSDF が事前に設計した戦略がどこで・どの程度ずれたかを、6 軸（GP / SI / FC / TC / DC / CSA）の事後診断として表現します。本フレームワークの完全な学術仕様は別途公開予定の論文でカバーされ、本書ではハイレベルの概念のみを提示します。

Appendix G 方法論上の前提

参照訳非依存（Reference-Translation Independence）

CATER v2 はモデル翻訳の存在を前提としません。原文・訳文・任意の Translation Brief・任意の Reference があれば評価が成立します。参照訳がある場合でも、それは経験的参照（Empirical Reference）の一形態として扱われ、評価そのものを駆動しません。

生成主体非依存（Producer Indifference）

翻訳が人手・NMT・LLM・ポストエディット・ハイブリッドのいずれで作られたかは、評価結果に影響しません。LLM プロンプト・スコアリングのどこにも生成主体は流れ込みません。

決定性（Determinism）

学術的再現性のため、すべての LLM 呼び出しは temperature = 0、top-p = 1 で実行されます。スコア計算式（EUC / REUC / RiskAdjustedER / AxisCap / GlobalGate）はコードで実行され、LLM には委譲されません。

簡易版スコアの扱い

簡易版（Basic）のスコアは LLM が直接判定した推定値です。論文や出版判定にはアドバンスド版（Advanced）の REUC / RiskAdjustedER パイプラインを通したスコアを使用してください。

公開可能範囲

本ページに掲載されている軸定義・スコア式・モード定義は、公開資料として配布可能な範囲です。Translation Strategy Design Framework（TSDF）の完全な学術仕様は別途公開予定の論文でカバーされます。

本 Appendix は CATER v2 — A Strategy-Theoretic Diagnostic Framework （Iida 2026-05-02）から、公開資料として配布可能な範囲を抜粋・再編集したものです。詳細な学術仕様は別途公開予定の論文を参照してください。Temperature = 0, top-p = 1 のもとで決定論的に実行され、スコア計算式はソースコードで実装されています。

CATER v2 — Reference-independent, producer-indifferent translation diagnostic. Reproduced at temperature 0.
Scoring math (REUC / RiskAdjustedER / AxisCap / GlobalGate) is computed deterministically by the orchestrator, not by the LLM.