

Fortress Translator 翻訳品質固定評価基準 v2.0

提示型・理解型を統合した汎用翻訳品質評価基準

【評価目的】

日本人旅行者が旅行現場で、訳文を相手へ提示する場合、または訳文を通じて相手の発言・文章を理解する場合に、原文の意味・必要情報・発話機能を忠実に保持し、対象言語として自然で読みやすく、想定用途において追加の修正・推測・再構成を原則として必要としない完成度に達しているかを評価する。

【利用形態】

区分	定義	評価上の重点
提示型	利用者が訳文を現地の相手へ見せる、送る、読み上げる。	修正せず直接提示できるか、相手が適切に対応できるか。
理解型	利用者が相手の発言・文章の訳文を読み、内容を理解して判断する。	再読・推測・再構成をせず、正確に即時理解できるか。

利用形態はテスト原文と翻訳方向・利用目的から採点前に固定し、候補ごとに変更しない。双方向通訳テストは各発話方向を個別に判定する。

【評価構造】

意味・関係 50 点、情報の保持 20 点、自然さ・文体 20 点、可読性 10 点の4軸、合計100点で絶対評価する。最終総合点は4軸得点の合計とする。

計算式

最終総合点 = 意味・関係 + 情報の保持 + 自然さ・文体 + 可読性

【評価設計の要点】

- 原文意味マップ、重要度、利用形態を採点前に固定する。
- 確認、依頼、拒否、許可、質問、報告等の発話機能を正式に評価する。
- 提示型は修正負荷、理解型は推測・再読・再構成の理解負荷で言語品質を判定する。
- 反実仮想修正テストにより、同一問題の重複減点を遮断する。
- AI一次採点および第三者監査では候補情報を伏せ、匿名の候補IDのみを用いる。

【基本原則】

- 点数を先に決めず、差異抽出から得点帯確定まで固定手順で処理する。
- 全候補へ同じ条文と同じ得点アンカーを適用する。
- 辞書差、好み差、自然な言い換えは減点しない。
- 同じ問題は最も直接関係する一軸だけで扱う。
- 名称、ブランド、作成元、既存評価、順位を採点根拠にしない。

【利用形態の判定境界】

状況	扱い
旅行者の日本語を現地語へ翻訳	原則として提示型。相手に見せる・伝える完成文として評価する。
現地語の発言・表示を日本語へ翻訳	原則として理解型。旅行者が正確に判断できる訳文として評価する。
文書作成・送信用の翻訳	提示型。送信前の修正が必要かを評価する。
双方向の会話・通訳	各発話方向を分離し、それぞれ提示型または理解型として採点する。

禁止

同一テスト内で、低品質な候補だけに有利となる利用形態へ変更してはならない。

【AI 一次採点の固定手順】

候補名、モデル名、作成元、既存評価、順位を伏せ、匿名の候補IDのみを用いる。各候補は必ず次の順序で処理する。

1. 利用形態の固定提示型または理解型をテスト単位で確定する。
2. 原文意味マップ主体、所有者、対象、行為、状態、条件、否定、時点、数値、場所、方向、発話機能、求める対応を分解する。
3. 重要度区分各意味単位を重大・中核・補助へ分類する。
4. 差異候補の全件抽出意味、情報、自然さ、可読性の差を点数入力前に列挙する。
5. 合理的な訳文理解対象言語の一般的な話者が当該場面で行う通常の理解を確定する。
6. 非減点差異の除外同義表現、必要な補完、数値同値、自然な省略、好み差を除外する。
7. 忠実度・実務影響証明意味・情報の差が、修正・確認・判断・処理へ与える影響を確定する。
8. 発話機能確認質問、確認、依頼、命令、拒否、許可、報告、推測の強さと方向を照合する。
9. 問題群の整理同一原因の反復と独立問題を区別する。
10. 反実仮想修正テスト一つの修正で消える複数の表面差を重複計上しない。
11. 一問題一軸各問題を意味、情報、自然さ、可読性の一軸へ所属させる。
12. 利用負荷の確定提示型は修正範囲、理解型は推測・再読・再構成範囲を確定する。
13. 得点帯・帯内整数重要度、範囲、回復可能性、利用負荷を固定アンカーへ照合する。
14. 整合検査減点理由、軸得点、総評、合計点を一致させる。

全候補確定後

同程度の重要度・利用負荷・処理差が、候補間およびテスト間で同じ得点帯に入っているかを確認した後に順位を算出する。

【原文意味マップ】

原文を、登場主体、所有者、対象、行為、状態、条件、否定、時点、数値、場所、連絡・受渡し方向、発話機能、求める対応へ分解する。原文の明示内容
と文脈から一意に導ける内容を区別する。

【重要度区分】

以下の項目は典型例であり、語の種類だけでは決めない。原文内の役割、誤差が判断・行動へ与える影響、回復可能性によって最終区分を確定する。

区分	典型例	採点上の扱い
A 重大	安全、医療、アレルギー、薬、法的手続き、重要な支払い・金額・日時・数量、本人確認、予約・識別番号、重要な禁止・否定。	改変・欠落は、原文内の役割、影響範囲、回復可能性に応じて該当軸へ大きく反映し、具体的な影響を減点理由へ記録する。
B 中核	行為者、対象、所有者、依頼・連絡・受渡し方向、場所、条件、状態、希望、確認、求める処理。	中心用件の忠実性を左右する。局所でも明確な変化は意味・情報へ反映。
C 補助	背景説明、非重要なニュアンス、丁寧さの細部、意味を変えない表現精度。	処理・命題が同じなら軽微減点または非減点。好み差は減点しない。

同じ時刻でも、雑談ではC、予約・服薬・搭乗ではAになり得る。場面別の固定語表ではなく、原文意味マップ上の役割で判断する。

【発話機能の固定確認】

原文の機能	変更として扱う例
可否・事実の確認	直接要求、命令、許可の付与、断定へ変わる。
依頼	可能性の説明、単なる提案、相手側の自由選択へ弱まる。
拒否・禁止	希望、推奨、可能性へ弱まる、または肯定へ反転する。
報告・申告	質問、推測、相手側の情報提示へ変わる。
許可・同意	義務、命令、単なる可能性へ変わる。
条件付き依頼	無条件要求、確定事項、別条件へ変わる。

非減点境界

丁寧さの調整や対象言語で自然な疑問形への変換で、発話の強さ・方向・読み手の対応が変わらない場合は減点しない。

【差異候補と採点対象問題】

区分	定義	扱い
差異候補	語句、構文、範囲、抽象度、明示性、順序、文体等に見つかった差。	発見時点では減点しない。
非減点差異	合理的な読み方で原文と同じ命題・必要情報・発話機能を保ち、用途上の修正・推測が不要な差。	問題数にも得点帯にも反映しない。
採点対象問題	意味忠実度、情報精度、対象言語品質、読み進めやすさのいずれかを実際に低下させる差。	一問題一軸で得点へ反映する。

【非減点差異フィルター】

確認項目	非減点	採点対象になり得る状態
命題・関係	行為者、対象、所有者、条件、否定、時点、確実度が同じ。	変化・曖昧化し、原文どおりの提示・理解に修正または確認が必要。
発話機能	質問・依頼・確認等の強さと方向が同じ。	確認が要求へ、依頼が許可へ、報告が質問へ変わる。
必要情報	識別・連絡・手続き・判断に必要な精度を保持。	対象、期限、能力、範囲、番号等が欠落・一般化・限定。
補完・省略	文脈から一意に導ける要素を自然に補完・省略。	新事実、人物属性、期限、義務、許可、選択肢を作る。
言語品質	同義で自然、統一的で、用途上の追加作業が不要。	文法、語彙、人称、属性、語調、構造に独立問題がある。

【忠実度・実務影響証明】

意味・関係または情報の保持の問題は、差の特定、合理的理解、因果・現実性を満たし、実務影響または忠実度影響のどちらかが成立した場合に採用する。

経路	成立条件	不成立
実務影響	案内、手続き、連絡、受渡し、判断、行動が現実的に変わる。	原文と同じ理解・同じ対応。
忠実度影響	処理が同じでも、命題・発話機能・必要情報を正確に提示・理解するには修正または確認が必要。	辞書範囲や好みだけの差。

実務影響の定型

訳文では____と理解されるため、原文なら____するところ、____する可能性がある。

忠実度影響の定型

訳文では____と理解され、原文の____が____へ変化・弱化する。原文どおりに使用するには____の修正または確認が必要である。

【反実仮想修正テスト | 重複減点の遮断】

- 意味・情報上の誤りを正しい内容へ置き換えたかと仮定する。
 - その修正だけで文法・語彙・構造の問題も消えるなら自然さへ重ねない。
 - 正しい内容へ直しても、語順、語調、人称、冗長さ等の独立問題が残る場合だけ自然さでも別問題とする。
 - 文法修正だけで意味も原文へ戻る場合は、問題の中心を意味・関係へ置く。
- 禁止
- 同じ語句を「意味が誤っている」「不自然でもある」という理由だけで二軸へ重複計上してはならない。

【提示型の修正負荷】

区分	定義	例
微調整	意味・使用可否を変えない任意の推敲。	より自然な語順、好みの語彙。
局所修正	一語・一短句・一構文を直せば解消する。	不適切な語、余分な助動表現。
節修正	一つの節または文を組み直す必要がある。	主客の混乱、破綻した接続。
全体修正	同じ原因が複数節へ広がり、出力全体で修正が必要。	話者属性の反復固定、文体全体の不統一。

【理解型の理解負荷】

区分	定義	例
負荷なし	一読で正確に理解でき、推測・再読が不要。	自然で明瞭な説明。
局所推測	一語・一短句を文脈から補えば理解できる。	不精密な語、軽い係り受け。
節再構成	一つの節または文を読み直し、関係を組み直す必要がある。	主客の混乱、接続破綻。
全体再構成	複数節の関係を推測し直さなければ理解が安定しない。	人物像・時系列・文体の広範な不統一。

【問題群と独立性】

一方だけを修正・再解釈しても他方が残る場合は独立問題とする。一つの修正方針または一つの推測で複数箇所が同時に解消する反復は一つの全体問題群とし、出現回数を機械的に加算しない。ただし範囲の広さは得点帯へ反映する。

異なる原文意味単位に属し、一方の修正または再解釈によって他方が解消しない、意味が成立しない、または一意に確定できない箇所は、それぞれ独立問題として全件記録する。同一原因による連鎖的な崩壊は一つの問題群として扱うが、影響する意味単位の数、重要度および範囲を得点帯へ反映する。

【利用完成度の分類】

区分	提示型	理解型	自然さ上限
A	修正不要で直接提示可能。	一読で正確に即時理解可能。	17-20
B	目立つ問題はあるが提示可能。	短い再読のみで理解可能。	16
C	一つの局所修正が必須。	一つの局所推測が必須。	15
D	全体修正一群または複数の局所修正が必須。	反復推測または一つの節再構成が必要。	12-14
E	多数の修正・節の書き直しが必要。	複数節または全体の再構成が必要。	0-11

【話者属性・人称・選択肢】

- **提示型**：未指定属性を一箇所て固定し利用者の変更が必要なら原則15点以下、複数節へ反復する全体修正は14点以下、未処理選択肢は13点以下、人物像の不統一は10点以下。
- **理解型**：未指定属性を事実として追加し人物理解を変える場合は意味・関係または情報の保持へ所属させる。人物理解を変えず、文章上の不統一だけが残る場合は自然さ・文体で扱う。

原文・テスト条件で属性が明示される場合、または対象言語で実質的に属性中立として通常使用される場合は上限を適用しない。

【1. 意味・関係：50 点】

命題、主客、行為者、対象、所有者、状態、条件、否定、可能性、確定度、時系列、因果、連絡・受渡し方向、発話機能を評価する。提示型・理解型で共通の基準を用いる。

得点帯	判定状態
50	意味・関係・発話機能を実質的に完全保持。意味修正・確認が不要。
49	C区分の微小な精度差が1件。中心用件は同じで正確に使用できる。
47-48	一つの局所的な意味修正、軽い発話機能変化、または複数の微小差。中核は容易に回復可能。
43-46	B区分の実質的な変化が1件、または局所的な意味変化が複数。確認・部分修正が必要。
36-42	重要な行為者・対象・所有者・条件・否定・方向の誤りが1件、または局所変化が広範。
25-35	独立した重要関係の誤りが2件以上、A区分の重大な変化、または用件の複数箇所が不安定。
0-24	中心目的、重要な否定・条件・主客が大きく崩れ、原文の用件を安定して伝達・理解できない。

帯内決定

上側は局所・容易に回復、中央は独立問題が複数または中程度、下側は中核へ波及・節修正または節再構成が必要な場合とする。

【2. 情報の保持：20 点】

原文に含まれる必要情報が、同じ対象・範囲・精度・形式で残っているかを評価する。意味関係が変わる場合は意味・関係を優先する。

得点帯	判定状態
20	A・B・Cの必要情報を実質的に完全保持。
19	C区分の詳細が1件弱体化・一般化。容易に回復可能。
18	C区分の差が複数、またはB区分の軽い精度低下が1件。
15-17	B区分の情報が一部欠落・追加・限定されるが、中心処理・理解は可能。
11-14	Aまたは重要なB情報の欠落・変更、あるいは複数情報の精度低下。
6-10	識別番号、条件、期限、対象情報等が大きく失われ、正確な対応・判断が難しい。
0-5	必要情報が広範に欠落・別物化し、復元が困難。

【数値・識別情報・固有名詞】

- 完全同値な数値・単位変換は非減点。
- 料金、時刻、数量、薬量、予約番号、荷物タグ、電話番号、URL等は、判断・照合への影響で重要度を決める。
- ハイフン・空白・文字種の形式差が照合を妨げない場合は非減点。
- 原文の誤記・不完全表記を外部知識で勝手に修正しない。

【3. 自然さ・文体：20 点】

対象言語として文法的・自然・明瞭で、人称、属性、敬称、語尾、語調、地域表現が一貫し、提示型では修正せず提示できるか、理解型では推測・再構成なしに理解できるかを評価する。

得点	判定状態
20	言語品質問題なし。提示型は完全に修正不要、理解型は一読で即時理解可能。
19	任意の微調整が1件。利用可能性への影響は極小。
18	軽微問題が1件または微小問題が複数。修正・推測なしで使用可能。
17	軽微問題が2-3件。違和感はあるが直接提示または一読理解が可能。
16	目立つ局所問題が1件。提示型は推敲推奨、理解型は短い再読が必要。
15	提示型は一つの局所修正、理解型は一つの局所推測が必須。
14	全体修正一群・二つの局所修正、または反復推測・一つの節再構成が必要。
12-13	全体修正に追加問題、三つ程度の独立修正、または複数の推測と節再構成が必要。
9-11	四つ以上の独立修正、人物像の不統一、未処理選択肢、または複数節の再構成が必要。
6-8	複数節の書き直し・全体的再構成が必要で、そのまま使用できない。
0-5	文法・語彙・文体が大きく崩れ、対象言語として安定しない。

固定運用

意味・情報へ所属した誤りを、出力全体が悪く見えるという理由で自然さへ重複計上しない。

【4. 可読性：10 点】

句読点、改行、文の区切り、反復、誤結合、極端な長文化、表示形式を確認し、スマートフォンで用件を順に追えるかを評価する。

得点帯	判定状態
10	区切りと構造が明瞭。提示・理解のどちらでも無理なく読み進められる。
9	局所的な区切り・接続問題が1件。
7-8	複数の読みづらさはあるが、用件を順に追える。
4-6	長文化、誤結合、区切り不足により再読・再構成が必要。
1-3	文の流れが大きく崩れ、内容の組み直しが必要。
0	文章構造が実質的に成立していない。

【不完全出力・選択肢残存】

状態	所属と扱い
途中終了による後続欠落	欠落情報は情報の保持。文の破綻が独立して残る場合のみ自然さまたは可読性。
全文が一段落	一段落だけでは非減点。追跡困難な場合のみ可読性。
属性・語尾の選択肢が残る	自然さ・文体。利用者の追加作業範囲を上限へ反映。
見出し・モデル名・注記が混入	用件理解を妨げる最も直接的な一軸で評価。
同一内容の重複	意味を増減させなければ可読性または自然さ。要求回数等が増えれば意味。

【原文の曖昧さ・矛盾・欠陥】

原文の状態	扱い
複数の合理的解釈	必要な解釈範囲を保つか確認。訳文が一つへ限定し対応・判断が変わる場合のみ評価。
原文自体が矛盾	理想原文へ勝手に修正しない。矛盾を含む用件を保持したか評価し、原文欠陥は別注記。
省略が対象言語で不自然	文脈から一意に導ける主語・目的語・接続を補完可。新事実の追加は評価対象。
誤記・不完全な固有名詞	与えられた表記を基準にし、外部知識で修正しない。

【汎用性条件】

特定の言語表現や一つのテストだけを根拠にした規則は一般条文にしない。条文は「利用形態 × 意味単位 × 重要度 × 合理的理解 × 利用負荷 × 実務影響」で説明できなければならない。

【重要な問題の表示方法】

安全、医療、金額、日時、数量、識別情報、行為者、連絡先、受渡し先などに関わる重要な問題も、該当する評価軸の得点へ反映し、問題の内容、重要度、影響経路を減点理由として全件記録する。4軸とは別の二値判定や追加減点は設けない。

【テスト結果の状態区分】

状態	定義	集計上の扱い
採点完了	翻訳結果が得られ、4軸得点を確定。	得点・平均・順位へ含める。
評価対象外 非対応	候補が対象言語・翻訳方向に対応していない。	0点にせず除外し、非対応として明示。
未測定 取得不能	一時的障害、アクセス不能等で翻訳品質を評価できない。	0点にせず未測定。取得条件に従い再取得する。
採点対象 不完全出力	入力を受理したが、空出力、途中終了、破綻文等を返した。	実際の出力として4軸採点する。

【単一テストと複数テストの順位】

- **単一テスト**：当該テストで採点完了した候補だけを順位付けする。
- **複数テスト総合**：事前に定めた共通テスト集合をすべて採点完了した候補だけを正式順位の対象とする。
- 一部テストが未測定・非対応の候補は、暫定平均と完了率を表示してもよいが、正式総合順位には入れない。
- 平均点と完了率を別々に表示し、未測定・非対応を0点として品質平均へ混入させない。
- 同点は同順位とし、同順位内の表示順は優劣を意味しない。

【減点理由・総評の固定記載】

- 得点へ使用した全問題を減点理由へ記載する。
- 意味・情報は、原文の意味単位、訳文の合理的理解、確定差、影響経路、所属軸・得点帯を記載する。
- 自然さは、対象言語上の問題、利用形態、修正負荷または理解負荷を記載する。
- 総評では新しい問題を追加せず、点数、主要問題、用途上の完成度をまとめる。
- 点数、減点理由、総評、順位が一致しない場合は採点無効。

【AI 一次採点の固定出力】

候補ID:

利用形態：提示型／理解型

意味・関係：__/50 情報の保持：__/20 自然さ・文体：__/20 可読性：__/10

最終総合点：__/100

減点理由：__

総評：__

【第三者監査】

第三者監査でも候補情報を伏せ、AI一次採点の総合点・4軸内訳、減点理由、総評を「翻訳品質評価シート」に基づいて5段階で監査する。

評価	判定	定義
5	妥当	点数・指摘・判断に異論なく修正不要。
4	概ね妥当	点数程度または軽微な判断差。
3	一部修正	点数または指摘に明確な修正が必要。
2	妥当性が低い	過大・過小評価、見落とし、誤認が複数。
1	妥当ではない	品質実態と大きく乖離し根本見直しが必要。

【最終原則】

評価対象はブランドではなく訳文である。利用形態を先に固定し、意味・情報は原文意味マップと発話機能、自然さは修正負荷または理解負荷、可読性は表示構造で評価する。同じ問題を重複させず、あらゆる旅行場面と翻訳方向で同じ論理が成立することを優先する。