JTF翻訳品質評価ガイドライン

第1版

（2018-11-15）

一般社団法人 日本翻訳連盟（JTF）

目次

目次 1

1. はじめに 4

1-1. 翻訳品質評価の課題と現状 4

1-2. 本ガイドラインの目的と意義 4

1-3. 本ガイドラインの適用範囲と利用者 4

2. 翻訳品質評価とは 5

2-1. 用語 5

品質（Quality） 5

品質評価（Quality evaluation/assessment） 5

仕様（Specification） 5

混同しやすい用語 5

2-2. 品質評価の目的 6

【補足】品質評価と品質保証 6

2-3. 品質評価の手法 6

A. 専門家による主観評価 7

B. 最終読者による主観評価 7

C. 評価メトリクスを用いた客観評価 7

エラーベースの評価メトリクスに関する注意点 7

D. その他の客観評価 8

a. ユーザビリティー評価 8

b. リーダビリティー評価 8

c. 法的要件などへの準拠 8

【補足】主観評価を活用する品質改善 8

【補足】機械翻訳における「自動評価」との違い 9

2-4. 品質評価で考慮する制約 9

3. JTF翻訳品質評価モデル 11

3-1. 本モデルの特徴 11

適用範囲 11

3-2. 標準的な評価フロー 11

1. 仕様を作成する 12

2. 評価者、評価対象、評価目的、評価手法を決める 12

3. 評価メトリクスを設定する 13

4. 評価メトリクスを使って評価作業を実施する 13

5. 評価結果を活用する 13

3-3. エラー・カテゴリー 14

A. 正確さ（Accuracy） 14

a. 誤訳（Mistranslation） 14

b. 抜けと余分（OmissionとAddition） 14

c. 未翻訳（Untranslated） 14
d. その他 14
B. 流暢さ（Fluency） 14
a. 誤入力（Typography） 15
b. 誤字（一部Spelling） 15
c. 同音異義語誤り（一部Spelling） 15
d. 文法誤り（Grammar） 15
e. 誤用（-） 15
f. コロケーション誤り（-） 15
g. 待遇表現誤り（一部Grammatical register） 15
h. 不整合（Inconsistency） 15
i. あいまい（Ambiguity） 15
j. 読解不能（Unintelligible） 16
k. その他 16
C. 用語（Terminology） 16
a. 指定用語違反（Inconsistent with termbase） 16
b. 特定分野用語違反（Inconsistent with domain） 16
c. 用語不統一（Inconsistent use of terminology） 16
d. その他 16
D. スタイル（Style） 16
a. 指定スタイル違反（Company styleとThird-party style） 17
b. 特定分野スタイル違反（一部Register） 17
c. スタイル不統一（Inconsistent style） 17
d. その他 17
E. 地域慣習（Locale convention） 17
a. 数値形式（Number format） 17
b. 日付形式（Date format） 17
c. 時刻形式（Time format） 17
d. 通貨形式（Currency format） 18
e. 度量衡形式（Measurement format） 18
f. 住所形式（Address format） 18
g. その他 18
F. デザイン（Design） 18
a. 全体デザインの問題（Overall design） 18
b. 局所フォーマットの問題（Local formatting） 18
c. 長さの問題（Length） 18
d. 文字切れ／はみ出し（Truncation/text expansion） 19
e. 非表示テキスト（Missing text） 19
f. マークアップ誤り（Markup） 19
g. 図表誤り（Graphics and tables） 19
h. その他 19
G. 事実性（Verity） 19
a. 当該地域で不適当（Locale-specific content） 19

b. 最終読者に不適当（End-user suitability） 19
c. 法的要件欠如（Legal requirements） 20
d. その他 20
H. その他 20
3-4. 重大度 20
深刻（Critical） 20
重度（Major） 20
軽度（Minor） 21
なし（None） 21
3-5. カテゴリー重み 21
3-6. 合否しきい値 23
目安となる合否しきい値 23
エラーの重複カウント 24
3-7. 評価シートのサンプル 24
付録 25
仕様のサンプル 26
ドキュメント・タイプ別のカテゴリー重み例 28
代表的な評価モデル 29
LISA QA Model 29
SAE J2450 29
MQM 29
DQF-MQM 30
ITS（Internationalization Tag Set）バージョン2.0 30
ISO 21999 30
ASTM WK46396およびWK54884 30
Garvinに基づく翻訳品質5分類 31
参考文献 32
関連情報 33
改訂履歴 34
日本翻訳連盟について 34
制作者について 34
著作権 35
利用許諾 35
免責 35
フィードバック 35
クレジット 36

1. はじめに

1-1. 翻訳品質評価の課題と現状

翻訳の「品質」は、何を指しているのか使う人によって異なることがある。そのため関係者どうしで共通認識を持つのが難しく、ビジネスがスムーズに進まないケースも見られる。

主観的な評価を避けようとして、一部の分野（ローカリゼーションなど）ではメトリクスによる品質評価手法が用いられることがある。しかしこのような手法は必ずしも日本国内の翻訳会社で広く採用されているとは言えない。また現在、欧州を中心とした海外で標準的な評価手法の作成が進んでいるものの、欧州言語を前提としている部分があるため、日本語翻訳で適用が難しい場面もある。

つまり、品質について関係者間で共通認識を持つことができ、さらに日本語翻訳も視野に入れた標準的な翻訳品質評価手法が望まれているのが現状であると言える。

1-2. 本ガイドラインの目的と意義

このような状況に対応する目的で、日本翻訳連盟（JTF）は業界団体として翻訳品質評価ガイドラインを提示する。本ガイドラインを活用することによって、受注者と発注者を中心とした関係者が共通認識を持ち、品質に関する合意を形成しやすくなるはずである。その結果としてビジネスが円滑化することが期待される。

また翻訳ビジネスはグローバルな取引が多いため、海外で開発されてきた評価手法との互換性にも配慮している。そのため海外企業とビジネスをする際にも齟齬や混乱を生じさせることなく利用できる。

1-3. 本ガイドラインの適用範囲と利用者

受発注者が存在する翻訳ビジネス場面での利用を想定している。また機械翻訳出力の人手評価や、実務翻訳教育における訳文評価で活用することも可能である。

翻訳ビジネスの場面においては、発注者となるクライアント企業や翻訳会社の担当者、受注者となる翻訳会社担当者や翻訳者を主な利用者として想定している。

2. 翻訳品質評価とは

本章では、翻訳品質に関する全般的な情報を解説する。次の章で取り上げる「JTF翻訳品質評価モデル」を利用する上で前提となる内容である。

2-1. 用語

まず本ガイドラインで用いる重要な用語を定義する。

品質（Quality）

翻訳成果物が、関係者間で事前に合意した仕様を満たす程度のこと。

翻訳サービスにおける「関係者」とは、発注者と受注者を指すことが多い。ただし状況によって異なり、たとえばトライアルにおける品質評価では発注者と受注者という関係はない。

品質評価（Quality evaluation/assessment）

翻訳成果物が仕様をどの程度満たしているか測定すること。

仕様（Specification）

クライアントや最終読者のニーズや目的などに基づき、翻訳成果物が持つべき要件をまとめたもの。文書化すると「仕様書」となる。

仕様に盛り込む項目は状況や翻訳案件で異なるが、納期、金額（単価）、分量、テキストに関する情報（例：言語方向、対象読者、翻訳の目的、言語使用域、専門分野）、参照資料（例：用語集、スタイルガイド、既存訳）、納品ファイル形式、使用ツール、品質保証や品質評価の方法（例：使用する評価メトリクス）などが挙げられる。仕様の文書サンプルを付録「仕様のサンプル」に示す。

日本の翻訳業界では手間がかかるため仕様書を作成しないこともある。しかしどのような翻訳を求めているのかを明示することで、受発注者で品質に対する共通認識を形成しやすくなる。また、後述するが、仕様は品質評価メトリクスを作る際の基礎にもなる。

混同しやすい用語

「品質評価」と混同して使われやすい言葉をいくつか紹介する。

- **品質保証（Quality assurance）**

仕様を満たすことを確実にし、実証するために行う活動全体。品質評価もその一環として実施される。

- **品質管理（Quality control）**
  品質保証を実現するための具体的な手段。
- **品質チェック（Quality check）**
  品質管理の手段の1つで、チェックシートやツールを用いることが多い。

2-2. 品質評価の目的

品質評価はさまざまな目的で実施する。本ガイドラインが主に想定している目的を示す。

- **納品物の検査**
  最終的に納品される成果物を検査するために実施される。ただし案件によって頻度（毎回、定期的）、範囲（全数検査、サンプル検査）、実施者（受注者側、発注者側、またはそのいずれかが委託した第三者）は異なることがある。結果を翻訳者や翻訳会社にフィードバックして改善を促すこともある。
- **成果物の中間検査**
  翻訳サービス提供者（例：翻訳会社やフリーランス翻訳者）がプロジェクトや作業の途中で、その時点での出来具合を測るために自身で実施する。
- **リソースの能力審査**
  発注者が取引相手（翻訳会社、フリーランス翻訳者）や社内担当者（翻訳者、チェッカーなど）の能力を審査するために、成果物を評価することがある。典型的にはトライアルである。

【補足】品質評価と品質保証

2-1で述べたように「品質評価」と「品質保証」は別の概念であるが、両者は関係している。品質評価は、品質保証という大きな活動の一部として実施することがある。最終納品物を検査してその結果を以降のプロジェクトに活かしたり、成果物を中間検査してその結果を納品に反映させたりする。つまり、品質評価は一度実施して終わりではなく、品質保証や品質改善のプロセスで繰り返し実施される。

2-3. 品質評価の手法

翻訳の品質はさまざまな方法で評価できる。以下では経営学者Garvinの分類（付録「Garvinに基づく品質5分類」）を参考にしつつ、直接的に翻訳成果物を評価する手法を挙げる。

これらのうち、どれを用いるのか、1つなのか組み合わせるのか、といった点で仕様作成時に関係者間で合意する。

A. 専門家による主観評価

専門家が自分の経験や感性に基づいて主観的に評価する手法。たとえば、評論家が小説や字幕の翻訳を評価したり、クライアント企業のエンジニアが技術的表現を評価したりするケースが該当する。産業翻訳では、後者のようなクライアント企業の専門家による主観評価がよく見られる。

B. 最終読者による主観評価

翻訳の最終読者（エンドユーザー）が自分の経験や感性に基づいて主観的に評価する手法。たとえば、企業サイトの「このページは分かりやすいですか？」といった質問に応じてユーザーが評価するケースが該当する。

C. 評価メトリクスを用いた客観評価

評価指標に基づいて定量的に評価する手法。「メトリクス」とは定量的な評価指標のことである。

特に欧米の翻訳業界では、エラーベースの評価メトリクスが用いられている。これは翻訳成果物に含まれるエラーに対し、重大度に応じて点数を付ける方法である。具体的にはLISA QA Model、SAE J2450、MQMなどがある（付録「代表的な評価モデル」参照）。

とりわけ複数の関係者（例：受発注者）がいる場合にこの手法は有用である。1-1で述べたように、関係者どうしで品質に関する共通認識を持つことは簡単ではない。しかしエラー項目と点数をあらかじめ明示することで、共通認識を持って合理的にコミュニケーションを図ることでビジネスをスムーズに進めやすくなる。次章で詳しく解説する本ガイドラインのモデルも、このエラーベースの評価メトリクスである。

エラーベースの評価メトリクスに関する注意点

エラーベースの評価メトリクス（本ガイドラインのモデルも含む）を用いる場合は以下のような点に注意が必要である。

- エラーベース手法は簡便に利用できるが、翻訳品質の一側面だけしか捉えていない点に常に注意を払うべきである。同手法では客観的検証が比較的容易なエラー項目を扱う。そのため、ある種の文書（例：広告）で重要となる主観的評価軸は備えていない。状況に応じてほかの評価手法（例：専門家や最終読者による主観評価）の利用を検討することが望ましい。

- 従来からエラーベース手法に対する批判がある。1つめは、文単位で評価すると「文章」に対する意識が薄くなりがちだという批判である。たとえば文章全体がある分野の慣習（例：学術論文、マニュアル）に従っているかを見落としがちになる。2つめは、固定的なエラー・カテゴリーを使った場合、さまざまな分野の文書に柔軟に対応できないという批判である。たとえば特許文書とマーケティング文書では、翻訳評価で重視される点が異なるはずである。
- メトリクスはいったん設定して終わりではなく、クライアントとの継続的な対話を通じて、エラー・カテゴリーやカテゴリー重みなどの充実化や改善を図ることが望ましい。主観的で暗黙的な評価基準を客観的なメトリクスに変換する努力である。

D. その他の客観評価

a. ユーザビリティー評価

ユーザビリティー指標を使って評価する手法。たとえば、和訳されたマニュアルを読んでユーザーが時間内に操作を完了できるかといった面から評価するケースが該当する。

b. リーダビリティー評価

リーダビリティー指標を使って評価する手法。たとえば、英訳した商品カタログをFlesch-Kincaid指標で評価するケースが該当する。

c. 法的要件などへの準拠

ある国で設けられている法的要件などを満たしているかで評価する手法。たとえば、ゲームがレーティングを取得できているかで評価するケースが該当する。

【補足】主観評価を活用する品質改善

評価が「主観」であると曖昧さがあるため避けるべきだと考えがちだが、品質を改善する源泉やヒントとなることが多い。重要なのは、主観評価の結果をいかに客観的なメトリクスに変換し、次に活かすかという点である。たとえば、ある翻訳成果物に対してクライアントが主観的な評価を下したとする（例：当社の製品カタログで敬体は読みにくい）。翻訳会社はクライアントと合意してこれをスタイル違反のエラー項目に加え、次回からメトリクスを使って評価できるようにする（図1参照）。これは品質保証活動とも関連している。

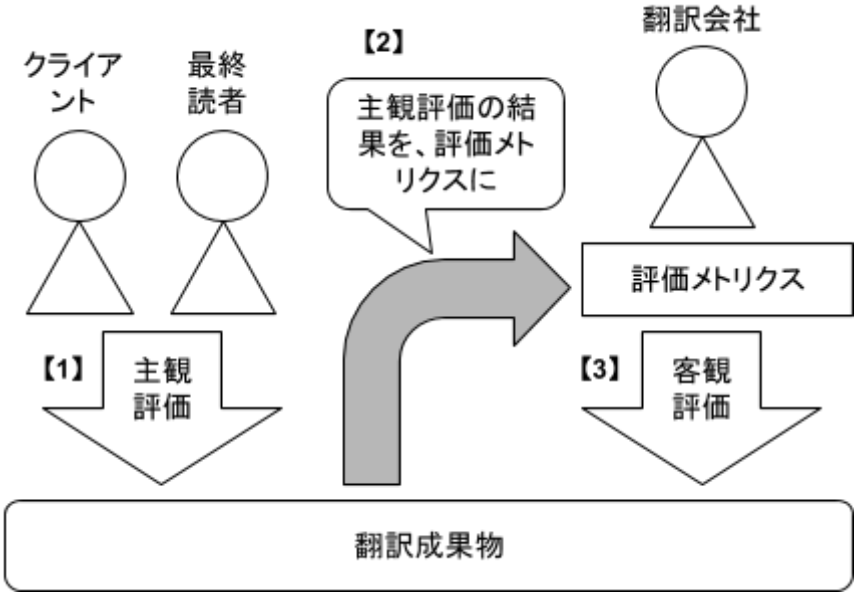

図1：主観評価結果のメトリクス化

【補足】機械翻訳における「自動評価」との違い

機械翻訳分野ではBLEUやMETEORといった尺度を用いた自動評価が行われることがある。これは基準となる参照訳とどのくらい近いかを計算して翻訳成果物を評価する方法である。対して2-3で紹介した手法は翻訳成果物そのものを測定するという点が異なる。つまり自動評価における参照訳自体も評価対象にできる。

2-4. 品質評価で考慮する制約

翻訳品質の評価で考慮しておくべき制約が存在する。ビジネスの場面においては、以下が代表的である。

- **納期**

  納期が短いと、訳文を十分に推敲できない、用語やスタイルを統一できないといった状況が生じることがある。納期が優先される場合、公正を期すために、たとえば一部のエラー・カテゴリーを適用しない、合否しきい値を通常より下げるといった合意を関係者間でし、仕様に盛り込む。
- **コスト**

  翻訳にかけられるコストや予算も制約となる。納期と同様に公正さを保てるような合意をし、仕様に盛り込む。
- **原文の質**

満足できる訳文に仕上がるかどうかは、原文の出来に左右される部分がある。原文の流暢さなどに問題があると判明した場合、訳文の合否しきい値を下げるといった対応も検討する。

3. JTF翻訳品質評価モデル

3-1. 本モデルの特徴

JTF翻訳品質評価モデルは、客観評価が可能となるエラーベースの評価メトリクスを利用しており、以下の特徴を持つ。

- 基本的には日本語への翻訳を評価する際に用いることを想定している。ただし英語への翻訳にはほぼそのまま適用可能である。また適切なエラー項目を用意すれば、それ以外の言語でも利用できる。
- 国際的に用いられているエラー・カテゴリーとの互換性を考慮しているため、海外企業と取引する場合でも大きな混乱を発生させることなく活用できる。
- 従来のエラーベース手法に対する批判に応えている。まず、文単位で評価するアプローチ（分析評価）と文章全体を評価するアプローチ（全体評価）の両方を設けており、どちらかまたは両方を用いる。また柔軟に評価メトリクスを設定できる点に加え、代表的な分野ごとに標準的なエラー・カテゴリーとカテゴリー重みも提示している。これにより、分野の実情に応じた柔軟な評価が簡便にできる。

前章の「エラーベースの評価メトリクスに関する注意点」で述べた通り、エラーベースの評価メトリクスは翻訳成果物の一側面しか見ていない。使用する際はエラーベース手法の限界に留意が必要である。

また本モデルはエラーベースの評価メトリクスであるが、主観評価手法を排除するわけではない。前章「主観評価を活用する品質改善」で触れたように、主観評価も併用して品質改善に取り組むことが望ましい。

適用範囲

また、本モデルが評価対象としているのは、情報の伝達を主目的とした産業翻訳（実務翻訳）である。情報伝達を主目的としない文芸など、メトリクスによる客観評価が馴染まない翻訳での利用は想定していない。

3-2. 標準的な評価フロー

本モデルを用いる際の標準的な評価フローを説明する。

1. 仕様を作成する

用語定義で述べたように、翻訳の品質とは「翻訳成果物が、関係者間で事前に合意した仕様を満たす程度のこと」である。そのため品質を評価するには、前提となる仕様の作成が不可欠である。とりわけ受発注者がいる場合は可能な限り仕様を明確にし、合意しておくことが望ましい。

2. 評価者、評価対象、評価目的、評価手法を決める

誰が、何を、何のために、どのように評価するのかという点である。仕様に基づいて決める。

- **評価者**

  納品物の検収が目的の場合、通常は翻訳者や翻訳チェッカーなど言語専門家が担当することになる。翻訳、多言語、専門知識などについての一定のスキルを持つ人が望ましい。また、最終読者にアンケートのような形で評価してもらうことも可能である。

  複数人で分担する場合、ばらつきが最小限になるよう、判断基準を設けたり事前トレーニングしたりすることが望ましい。
- **評価対象**

  通常は原文が存在する「訳文」が評価対象となる。ただし「原文」そのものを評価対象としてもよい（その際「正確さ」は測定できない）。
- **評価目的**

  受発注者がいるケースでは「納品物の検査」が基本的な目的となるが、翻訳会社が社内で「成果物の中間検査」や「リソースの能力審査」を目的に実施してもよい。
- **評価手法**

  このフローでは「評価メトリクスを用いた客観評価」である本モデルを用いることを想定している。ただし「2-3. 品質評価の手法」で説明した手法から、翻訳成果物の特徴やニーズに合致したものを追加してもよい。たとえば納品物をクライアント側で評価するケースであれば「専門家による主観評価」を追加で実施してもよい。

  本モデルでは従来と同様に1文ごとの評価を想定しているものの、文章全体も評価対象にできる点に特徴がある。そのため、以下のどちらか一方または両方の評価アプローチを用いるのかも決定する。

  - **分析評価**

1文（またはCATツールの1セグメント）ごとに評価する。全体評価と比較してある程度の時間がかかる。エラーベース手法でよく用いられる。全数検査することもサンプル検査することもできる。

- **全体評価**

  文章全体をまとめて評価する。最終読者など、専門家以外が評価する際も負担が少ないことが利点である。分析評価で用いるエラー・カテゴリーや重大度はそのままでは使いにくいことがあるため注意が必要である。
  たとえば、最終読者がウェブサイトの日本語訳全体について「流暢さ」のみを評価するケースが考えられる。

3. 評価メトリクスを設定する

仕様を参照しつつ、評価メトリクスを構成する以下の要素について設定する。

- **エラー・カテゴリー（3-3参照）**
- **重大度（3-4参照）**
- **カテゴリー重み（3-5参照）**
- **合否しきい値（3-6参照）**

たとえばマーケティング資料の翻訳であれば、エラー・カテゴリーで「正確さ」、「流暢さ」、「スタイル」の3つを使用し、そのうち「スタイル」を重視してカテゴリー重みを1.5倍にするといった設定である。特別な事情がなければ、デフォルトを使用しても構わない。

4. 評価メトリクスを使って評価作業を実施する

「分析評価」を用いる場合、1文ずつ評価作業を実施する。設定したエラー・カテゴリーに該当するエラーが翻訳成果物の中に見つかると、それに対してカテゴリー重みと重大度に応じて点数を付ける。エラーと点数を記録する評価シートは「3-7. 評価シートのサンプル」を参考にして自作したり、翻訳支援ツールに備わっている機能を利用したりする。

一方「全体評価」を用いる場合、文章全体をまとめて評価する。分析評価と異なり、評価シートはシンプルなものとなる。

5. 評価結果を活用する

評価結果が得られたら、納品物の合否判断を下す、リソースの能力を判断する、品質改善に役立てるといった場面で活用する。

3-3. エラー・カテゴリー

本モデルでは以下のエラー・カテゴリーを設けており、概要レベルと詳細レベルの2レベルがある。どのレベルと項目を用いるかは、分野や案件の状況に応じて取捨選択できる。通常の評価であれば概要レベルのみ、より厳密に評価したい場合は詳細レベルも用いる。またカテゴリーは、クライアントの要望などに応じて独自に設けてもよい。

以下では、概要レベルの項目を見出しとしてエラー・カテゴリーを説明する。参考までにMQM（Multidimensional Quality Metrics）の項目との対応関係をカッコ内に示す。

A. 正確さ（Accuracy）

原文の情報や意味が訳文に過不足なく盛り込まれているかの程度。対訳での確認が必要となる。詳細レベルの項目は以下の通り：

a. 誤訳（Mistranslation）

原文と訳文が伝える情報や意味にずれがある。例：英語原文では否定なのに、日本語訳文は肯定になっている。

b. 抜けと余分（OmissionとAddition）

原文に対応するテキストが訳文にない、または対応しないテキストがある。意味的というより、外見上の抜けや余分。例：原文にある段落が訳文にない。

c. 未翻訳（Untranslated）

翻訳すべきテキストが原文のまま残っている。例：英日翻訳で「I will 学校に行く」と英語が残っている。

d. その他

正確さに分類するのが妥当なエラーで、上記のいずれにも該当しないもの。

B. 流暢さ（Fluency）

あるテキストが形式として整っているかの程度。主観的な読みやすさというより、文法など形式面に注目する（well-formedness）。訳文のみならず原文の評価にも使える。詳細レベルの項目は以下の通り：

a. 誤入力（Typography）

入力時に不注意で発生したと思われる誤りがある。例：「私はは昨日」。

b. 誤字（一部Spelling）

ひらがな、カタカナ、漢字、句読点の表記や、欧文のスペルに誤りがある。例：「予定どうり」、「設定しずらい」、「スエーデン」、「二本語」、「commitee」。

c. 同音異義語誤り（一部Spelling）

同音異義語の誤りがある。例：「意外」を「以外」、「保証」を「補償」、「初めて」を「始めて」。

d. 文法誤り（Grammar）

文法上の誤りがある。例：日本語訳文で助詞「が」を使うべき部分で「は」となっている。

e. 誤用（-）

本来の意味とは違う意味で使われている。例：「煮詰まる」を行き詰まる意味、「おもむろ」を突然の意味で使っている。

f. コロケーション誤り（-）

コロケーション（語どうしの慣用的なつながり）に誤りがある。例：「約束を保つ」

g. 待遇表現誤り（一部Grammatical register）

敬語などの待遇表現が文脈上適切ではない。例：謙譲語を用いるのが適当なのに使っていない。

h. 不整合（Inconsistency）

同一のテキスト内に、記述の不整合がある。例：ある場所で「2015年に……」とあるのに、別の場所で「同じ2016年に……」とある。

i. あいまい（Ambiguity）

係り受けなどが原因で、さまざまに解釈可能されうる部分がある。例：「若い先生の犬」

j. 読解不能（Unintelligible）

言葉として読めない部分がある。例：「私は□□□□□」。

k. その他

流暢さに分類するのが妥当なエラーで、上記のいずれにも該当しないもの。

C. 用語（Terminology）

言葉づかい一般ではなく、ある分野内、企業内、または製品内で用いられる専門用語に関する問題。産業翻訳では、用語を的確かつ一貫して使うことが求められることが多い。詳細レベルの項目は以下の通り：

a. 指定用語違反（Inconsistent with termbase）

指定された用語が使われていない。例：用語集で「Configuration」を「構成」としているのに「設定」と和訳している。

b. 特定分野用語違反（Inconsistent with domain）

指定用語ではないものの、ある専門分野で一般的な用語が使われていない。例：自然言語処理分野の「named entity」に対して「固有表現」ではなく「名前付きエンティティー」が使われている。

c. 用語不統一（Inconsistent use of terminology）

指定用語ではないものの、用語が統一されていない。例：「戻り値」と「返り値」の両方、「犬」と「イヌ」の両方がある。

d. その他

用語に分類するのが妥当なエラーで、上記のいずれにも該当しないもの。

D. スタイル（Style）

表記のスタイルに関する問題。産業翻訳では、複数人が関与したり既存訳と整合を図ったりすることが多いため、スタイルの統一は重要となる。詳細レベルの項目は以下の通り：

a. 指定スタイル違反（Company styleとThird-party style）

指定されたスタイルが使われていない。例：スタイルガイドで常体（である調）と指定があるのに敬体（ですます調）を使っている。

b. 特定分野スタイル違反（一部Register）

指定スタイルではないものの、ある専門分野で一般的なスタイルが用いられていない。例：学術論文で敬体（ですます調）が使われている。

c. スタイル不統一（Inconsistent style）

指定スタイルではないものの、スタイルが統一されていない。例：箇条書きに常体と敬体の項目が混在、「受入テスト」と「受け入れテスト」が混在している。

d. その他

スタイルに分類するのが妥当なエラーで、上記のいずれにも該当しないもの。

E. 地域慣習（Locale convention）

言葉としては誤りではないが、訳文が読まれる地域の慣習に合致していない問題。詳細レベルの項目は以下の通り：

a. 数値形式（Number format）

数値の表現に問題がある。例：独日翻訳で小数点が「9,4」のようにカンマのままである。

b. 日付形式（Date format）

日付の表現に問題がある。例：英国向けパンフレットに「Jan 21, 2017」と米国式で記載されている。

c. 時刻形式（Time format）

時刻の表現に問題がある。例：24時間表記が一般的でない地域で24時間表記になっている。

d. 通貨形式（Currency format）

通貨の表現に問題がある。例：日本向けに仏日翻訳されたカタログの通貨がユーロのままである。

e. 度量衡形式（Measurement format）

度量衡の表現に問題がある。例：日本向けに英日翻訳された製品案内でヤード・ポンド法が使われている。

f. 住所形式（Address format）

住所の表現に問題がある。例：米国向けに日英翻訳された会社案内に日本の郵便マーク（〒）が使われている。

g. その他

地域慣習に分類するのが妥当なエラーで、上記のいずれにも該当しないもの。

F. デザイン（Design）

テキストの見た目に関する問題。ただし翻訳段階ではなく、DTPなどの編集段階での対応となることがある。詳細レベルの項目は以下の通り：

a. 全体デザインの問題（Overall design）

色、フォント、ページ区切りなど全体的なデザインに問題があったり、レイアウト崩れが見られたりする。例：原文では1ページ内にテキストが収まっていたが、訳文では2ページ目にはみ出している。

b. 局所フォーマットの問題（Local formatting）

フォント、字間、行間、インデントなど、全体ではなく局所で見た目上の問題がある。例：ある段落だけ字下げされていない。

c. 長さの問題（Length）

原文と比較し、訳文テキストが長すぎるまたは短すぎる。例：箇条書きの1項目だけが長すぎて違和感を覚えさせる。

d. 文字切れ／はみ出し（Truncation/text expansion）

スペースが足りないため、訳文テキストが切れたりはみ出したりしている。例：ボタンの英語原文が「Download」だが、日本語訳文は「ダウンロー」としか表示されていない。

e. 非表示テキスト（Missing text）

訳文テキストが存在するのに表示されていない。例：スライド上で吹き出しが拡大し、文字がその背後に隠れている。

f. マークアップ誤り（Markup）

原文と比較し、HTMLなどのマークアップが抜けている、余分にある、または位置が異なる。例：英語原文「Click the File button.」に対し日本語訳が「ファイルボタンをクリックします」となっている。

g. 図表誤り（Graphics and tables）

図や表のフォーマットに関する問題がある。例：図のコールアウトの位置が原文と違う。

h. その他

デザインに分類するのが妥当なエラーで、上記のいずれにも該当しないもの。

G. 事実性（Verity）

訳文の情報に、事実や現実世界との齟齬があるかどうかに関する問題。詳細レベルの項目は以下の通り：

a. 当該地域で不適当（Locale-specific content）

原文が読まれる地域（国）では正しい情報であるが、訳文が読まれる地域では不適当な情報となる。例：米国で書かれた英語パンフレットに無料通話の電話番号が掲載されているが、それが和訳されて日本国内からかけたら通じない。入力フォームに「州」の選択肢があるが、日本の都道府県が選択できない。

b. 最終読者に不適当（End-user suitability）

訳文の最終読者にとって適当ではない情報がある。そもそもの原文に問題があることが多い。例：マニュアルの指示に従うと機器が破損し得る。

c. 法的要件欠如（Legal requirements）

仕様に盛り込まれている法的要件を訳文テキストが満たしていない。例：日英翻訳したゲーム内の表現が問題となり、米国のレーティングを取得できない。

d. その他

事実性に分類するのが妥当なエラーで、上記のいずれにも該当しないもの。

H. その他

上記のカテゴリーに入らないエラー。

上記のうち、純粋に言語表現が関わるのはA～Eまでである（特にAとBは機械翻訳の人手評価で中心となる）。またFやGはテキストの外部にある「見た目」や「現実世界」に関係する。機械翻訳システムが扱えない部分であり、人間の判断が必要となる。ちなみにA以外は対訳が不要で、一言語のテキスト（例：原文のみ）を評価するのにも使える。

3-4. 重大度

あるエラーがどのくらい重いのかを示す程度。重大かどうかは翻訳の目的（例：マニュアルなら機器を操作ができる、広告なら顧客を獲得できる）を達成できるかどうかという視点から判断する。

各重大度の判断目安と標準の点数[1]を、高い順に以下に示す。標準点数は、関係者どうしの合意で調整を加えてもよい。

深刻（Critical）

標準点数：100

翻訳成果物の使用が不適当となるエラー。その訳文を読んだ結果、健康被害、経済的損失、社会的な評価毀損などをもたらす可能性がある。深刻なエラーは公開前に修正が必要である。

例：機器説明書の英日翻訳でnotを訳し漏らしたため、ユーザーが危険な操作を実行する。

[1] LISA QA Modelなど従来のモデルでは、標準点数を高い順に10、5、1とするケースがある。しかし本当に不合格にすべきエラー（例：訳文通り使うとユーザーが怪我をする）がある場合、深刻が10点だとしきい値を超えずに合格してしまうことがある。そのため本モデルでは深刻を100点として差を付けている。

重度（Major）

標準点数：10

翻訳成果物の理解しやすさや使いやすさに影響を与えうるエラー。翻訳成果物の使用が不適当とまでは行かないが、読者を混乱させる可能性がある。重度のエラーは公開前に修正するのが望ましい。

例：和訳で「指示する」とすべきところが「支持する」となっている。

軽度（Minor）

標準点数：1

翻訳成果物の理解しやすさや使いやすさに大きな影響を与えないエラー。読者は自分自身で補完して理解できる。軽度なエラーの修正は必ずしも必要ではない。

例：数字は半角にすべきとスタイルに指定があるが、全角が用いられている。

なし（None）

標準点数：0

エラーとして数えないもの。以下のような場合に用いる：

- 原文の誤りから発生したエラーで、翻訳者に責任がないもの
- 翻訳者に提供していない参考資料（例：用語集、スタイルガイド、既存訳、社内情報）に基づくエラー
- 評価担当者の好みによる修正（いわゆるpreferential error）。「専門家による主観評価」を併せて実施するケースで使用
- 一括修正できるような繰り返しのエラーで、2回目以降のもの（例：「iOS」をすべて「IOS」と誤字）。
    - ただしエラーの重複カウント（後述）をする場合は「なし」としない

3-5. カテゴリー重み

分野やドキュメント・タイプによって、どのエラー・カテゴリーを重視するのかが異なることがある。そのため関係者どうしの合意でエラー・カテゴリーの重み付けを変更できる。重み付けを大きくしたカテゴリーは、より合計点への影響が大きくなる。

本評価モデルでは、4段階の重みを想定している。各段階の名称と推奨の重み点数は以下となる。

重み名	点数
とても重視	2.0
やや重視	1.5
普通	1.0（標準）
あまり重視しない	0.5

特に重み付けを調整しなければ、すべてのカテゴリーで標準の「1.0」にする。重みを調整する場合、本モデルでは平均で1.0になる運用を想定している。たとえば4つカテゴリーがあり、重みを「1.5」、「1.0」、「1.0」、「0.5」としたら、平均で1.0になる。何かを0.5上げたら何かを0.5下げるといった対応が簡単である[2]。

ここで、重みを調整した例を2つ挙げる。特許文書で「正確さ」と「用語」を重視しつつ、「地域慣習」と「デザイン」は評価対象外としたケースである。

カテゴリー	重み	カテゴリー	重み
正確さ	2.0	スタイル	0.5
流暢さ	1.0	事実性	0.5
用語	1.5	その他	0.5

また海外の消費者向け製品パンフレット[3]で「流暢さ」、「地域慣習」、「デザイン」を重視するケースでは以下のようになる。

カテゴリー	重み	カテゴリー	重み
正確さ	0.5	地域慣習	1.5
流暢さ	1.5	デザイン	1.5

[2] もし「1.5」、「1.0」、「1.0」、「1.0」のように平均が1.0にならない値を設定したら、重みを相対化して調整することになる。「相対重み ＝ 重み付け設定点数 ÷ カテゴリー合計設定点数 × 使用カテゴリー個数」の式で算出し、それぞれ「1.33」（1.5 ÷ (1.5+1.0+1.0+1.0) × 4）、「0.89」（1.0 ÷ (1.5+1.0+1.0+1.0) × 4）、「0.89」（同上）、「0.89」（同上）となる。なお、JTFが提供する「評価シートサンプル」ではこの計算が自動でできる。

[3] 繰り返しとなるが、メトリクスによる評価は簡便であるものの、品質の一側面しか見ていない。たとえば消費者向けパンフレットでは「最終読者による主観評価」を併用した方が妥当な可能性もある。ドキュメント・タイプや状況に合わせて評価方法を採用すべきである。

用語	0.5	事実性	1.0
スタイル	1.0	その他	0.5

付録「ドキュメント・タイプ別のカテゴリー重み例」で、さまざまなドキュメント・タイプにおける重み付けの例を紹介している。各分野における専門家の意見を取り入れて作成しているため、重み付けを設定する際の参考にしていただきたい。

3-6. 合否しきい値

もし成果物に対して合否判定をする場合、合計の点数がしきい値を超えていたら不合格の判断を下すことになる。そのため関係者間であらかじめ合否しきい値を決め、合意しておく必要がある。

「分析評価」では1文（1セグメント）ずつ確認し、見つかったエラーについて、1件ごとに以下のようにエラー点数を付ける。

エラー点数 = カテゴリー重み × 重大度点数

対象すべてで評価作業が完了したら、このエラー点数をすべて足し、合計エラー点数を算出する。一般的に、エラー点数は標準化（英語からの翻訳では1,000ワードあたりなど）して算出することが多い。このエラー点数がしきい値内にあれば、合格という判断をする。

本モデルでも1,000単位（ワードまたは文字）あたりへの換算を想定しており、以下の式で計算する。

1,000単位あたり点数 = 合計のエラー点数 ÷ 評価対象ワード（文字）数 × 1,000

たとえば合計のエラー点数が23で、2,500ワードの翻訳成果物を評価したとすると、「23 ÷ 2,500 × 1,000」で「9.2」となる。もしこの数字が指定のしきい値よりも高ければ不合格、低ければ合格という判定を下せる。

目安となる合否しきい値

本モデルでは「分析評価」の場合、以下を目安となる合否しきい値としている。実際には、案件や分野による特徴、制約（例：納期、コスト、原文の質）を加味してしきい値を修正して利用する。

言語方向	目安となる合否しきい値[4]
英語 → 日本語	50点（原文1,000ワードあたり）
日本語 → 英語	35点（原文1,000文字あたり）

一方「全体評価」では、評価対象全体に対し、各カテゴリーで1回だけ点数を付けることになる。全体評価における目安の合否しきい値については示さない。

エラーの重複カウント

実際にエラーを見つけてカウントする際、同じエラーを重複してカウントすべきかどうかで判断に迷うことがある。たとえば「アクセシビリティー」とすべき複数の箇所で「アクセシビリティ」となっているケースである。重複カウントは合否に影響することがあるため、事前に考慮しておく必要がある。

本ガイドラインでは、重複カウントするかどうかは関係者どうしの合意で決定することを原則にしている。

ただし上記の2つの例のように、「軽度なエラーで一括置換が可能な場合」には、重複カウントせずに1回と数える考え方がある。また合意によっては、軽度なエラーで一括置換が可能であっても、一定個数（例：10）より多ければ「重度」のエラー1回と数える運用方法もある。

3-7. 評価シートのサンプル

評価時に用いる評価シートは、スプレッドシートで作成したり、翻訳支援ツールに搭載の機能を使ったりする。以下にスプレッドシート形式の評価シートのサンプルを提示する。英日翻訳の成果物を評価した例である。

[4] この目安は、翻訳者の1日あたりの作業で深刻エラー1つ相当の点数（100点）を超えたら不合格という考え方をベースにしている。一般的に、翻訳者の1日あたりの作業量は英日で2,000ワード、英日で3,000文字（ともに原文ベース）程度とされる。つまり、英日は2,000ワードで100点なので1,000ワード換算で「50点」、日英は3,000文字で100点なので1,000文字換算で「33.3 ≒ 35点」とした。ただしこの目安は暫定的なもので、将来的に利用者の意見などを聞いて修正する。

No.	ファイル名	位置	原文	訳文	カテゴリー	重み	重大度	重大度点数	エラー点数	評価者コメント
1	abc.docx	p. 3	I have a pen.	私はペンを所有する。	スタイル	1.0	軽度	1	1	敬体にする。
2	abc.docx	p. 4	I have an Apple.	私はりんごを所有しています。	正確さ	1.5	重度	10	15	ここは「アップル製品」。
3	abc.docx	p. 7	No objection!	意義なし！	流暢さ	1.0	重度	10	10	「異議」にする。
4	abc.docx	p. 8	However,…	しかれども、…	流暢さ	1.0	なし	0	0	表現が硬いので「しかし」に。
5	…	…	…	…	…	…	…	…	…	…

上記のサンプルの場合、エラー点数の合計は26（1 + 15 + 10 + 0）となる（なお0点部分は評価者の好みによる修正なので「なし」）。原文ワードが2,500だった場合、1,000ワードあたりに換算すると「26 ÷ 2,500 × 1,000」で「10.4」となる。もし合否判定を実施し、事前に合否しきい値を50と設定していた場合、このサンプルの納品物は合格という判定が可能となる。

付録

仕様のサンプル

仕様のサンプルを示す。翻訳会社とクライアント企業との間で作成されたことを想定している。項目は一例であり、実際には受発注者との間で追加または削除することになる。本ガイドラインでは評価実施が前提であるため「品質評価方法」という項目を入れている。

翻訳仕様書

株式会社ABCD

2018年9月1日

以下の条件で翻訳をお願いします。

プロジェクト関連情報	
納期	2018年10月15日 18:00
金額（単価）	25円／ワード（マッチ率によるスライドなし）
分量	3,532ワード
納品ファイル形式	pptxファイル
使用ツール	特に指定なし
品質保証	翻訳会社内でバイリンガルチェックを実施
テキスト関連情報	
言語方向	英日
対象読者	弊社スマートスピーカー製品を販売したい小売店
翻訳の目的	製品を購入してもらう
言語使用域	敬体を使う。ただしカジュアル過ぎない表現で
専門分野	IT、ビジネス
ドキュメント種類	製品紹介のプレゼン資料
参照資料	
用語集	なし

<table>
<tr><td>スタイルガイド</td><td>JTF日本語標準スタイルガイド</td></tr>
<tr><td>翻訳メモリー</td><td>なし</td></tr>
<tr><td colspan="2">品質評価方法</td></tr>
<tr><td>評価目的</td><td>納品物の検収</td></tr>
<tr><td>評価者</td><td>発注者（株式会社ABCD）側のレビュアー</td></tr>
<tr><td>評価手法</td><td>分析評価で、全数検査</td></tr>
<tr><td>エラー・カテゴリーと重み</td><td>以下のカテゴリーと重みを使用する：
<table>
<tr><th>カテゴリー</th><th>重み</th></tr>
<tr><td>正確さ</td><td>1.0</td></tr>
<tr><td>流暢さ</td><td>1.5</td></tr>
<tr><td>スタイル</td><td>0.5</td></tr>
<tr><td>デザイン</td><td>1.5</td></tr>
<tr><td>事実性</td><td>1.0</td></tr>
<tr><td>その他</td><td>0.5</td></tr>
</table></td></tr>
<tr><td>重大度</td><td>以下の重大度と点数を使用する：
<table>
<tr><th>重大度</th><th>点数</th></tr>
<tr><td>深刻</td><td>100</td></tr>
<tr><td>重度</td><td>10</td></tr>
<tr><td>軽度</td><td>1</td></tr>
<tr><td>なし</td><td>0</td></tr>
</table></td></tr>
<tr><td>合否しきい値</td><td>原文1,000ワードあたり50点未満とする</td></tr>
<tr><td>不合格時の対応</td><td>翻訳会社にフィードバックする</td></tr>
</table>

ドキュメント・タイプ別のカテゴリー重み例

さまざまなドキュメント・タイプ別にカテゴリーの重み付け例を示す。各分野における複数の専門家の意見に基づく数字[5]ではあるが、本ガイドラインが定めるルールではない。あくまでメトリクスを設定する際の参考値として活用していただきたい。

なお、ここでは「その他」は除いてある。また評価ですべてのカテゴリーを用いる必要はなく、どれを使うかは関係者間で合意する。

ドキュメント・タイプ	正確さ	流暢さ	用語	スタイル	地域慣習	デザイン	事実性	説明
特許明細書	2.0	1.0	1.5	1.0	0.5	0.5	0.5	原文情報が伝わるよう「正確さ」と、分野特有の「用語」がきちんと使われていることを重視。
マニュアル、技術文書（仕様書など）	2.0	0.5	1.0	1.0	1.0	0.5	1.0	原文の情報が正しく伝わるよう「正確さ」を最重視。マニュアルでは、機器やソフトウェアのUIで使われる「用語」と合致しているかを重視したり、問題なく動作するか実機チェックして「事実性」を重く見たりすることもある。
学術論文	2.0	1.0	1.0	1.0	0.5	0.5	1.0	原文の情報が正しく伝わるよう「正確さ」を最重視。場合によっては内容的な正しさをチェックし、「事実性」に重みを付けることもある。
決算書類、財務諸表	2.0	0.5	1.5	0.5	1.0	0.5	1.0	原文情報が伝わっているかを「正確さ」で、分野特有の用語が使われているかを「用語」で重視する。内容的な正しさを「事実性」で確認することもある。
治験実施計画書、医療診断書、医薬品添付文書、MSDS／SDS	2.0	0.5	1.5	1.0	0.5	0.5	1.0	原文情報が伝わっているかを「正確さ」で、分野特有の用語が使われているかを「用語」で重視する。内容的な正しさを「事実性」で確認することもある。
契約書、法令文書、規格、社内規定、金融関連文書（経済・市場分析など）	1.5	1.0	1.0	1.0	1.0	0.5	1.0	原文情報が伝わるよう「正確さ」を重視。内容的な正しさを確認する場合に「事実性」を重くすることも。
報告書、白書、議事録	2.0	1.0	1.0	0.5	0.5	1.0	1.0	原文情報が伝わるよう「正確さ」を重視。

[5] このカテゴリー重み付け例は、意見を取り入れて更新していく予定である。

ユーザー・インターフェイス	1.0	0.5	1.5	1.0	1.0	1.0	1.0	ボタン名などが統一されていることが重要であるため「用語」を重視する。ソフトウェア画面上での文字切れやはみ出しを「デザイン」で、実際にソフトウェアが動作するかを「事実性」で見ることもある。
マーケティング資料、広告、プレスリリース、企業ホームページ、会社案内、雑誌記事など	1.0	1.5	0.5	1.0	1.0	1.0	1.0	最終読者が受ける印象を考慮するため、「流暢さ」を最重視し、場合によっては「スタイル」も重く見る。一方、広告などクリエイティブなドキュメントの場合に「正確さ」や「用語」はあまり重視しないこともある。

代表的な評価モデル

業界で用いられている代表的な評価モデルには以下がある。簡単な解説を加えて紹介する。

LISA QA Model

ローカリゼーションの業界団体であるLISA（Localization Industry Standards Association）が開発したエラーベースの評価手法。評価方法自体と評価用ソフトウェアの両方を指す。

エラー・カテゴリーには、誤訳（Mistranslation）、正確さ（Accuracy）、用語（Terminology）、スタイル（Style）などの7つがある。また重大度は深刻（Critical）、重度（Major）、軽度（Minor）の3段階があり、それぞれ10点、5点、1点が割り当てられている。

1995年にバージョン1が公開されたが、2011年にLISAが解散したため、バージョン3.1からは更新されていない。

SAE J2450

自動車や航空機に関する業界団体であるSAE（Society of Automotive Engineers）による標準で、2001年に発行された。自動車業界でよく用いられているエラーベースの評価手法である。

エラー・カテゴリーには、用語間違い（Wrong Term）や意味間違い（Wrong Meaning）など7つがある。また、重大度は各カテゴリーで深刻（Serious）と軽度（Minor）の2段階が設けられている。

MQM

MQM（Multidimensional Quality Metrics）は、欧州委員会（EC）が資金提供しているQTLauchPadプロジェクトとドイツ人工知能研究センター（DFKI）が開発し、2015年にバージョン1.0が公開された。

MQMでは、LISA QA Modelなどで従来から用いられてきた文レベルでの評価（分析評価）に加え、文章全体レベルでの評価（全体評価）も提唱している。また、ドキュメント・タイプに応じて柔軟にエラー・カテゴリーを選択してメトリクスを作成できる仕組みを取り入れている。

エラー・カテゴリーには、正確さ（Accuracy）、流暢さ（Fluency）、用語（Terminology）、スタイル（Style）など上位レベルで7つがある。重大度は、深刻（Critical）、重度（Major）、軽度（Minor）、なし（None）の4つを設け、それぞれデフォルトの点数として100、10、1、0を設定している。

DQF-MQM

翻訳関連の組織であるTAUS（Transmation Automation User Society）が提供するDQF（Dynamic Quality Framework）のエラー・カテゴリーと、MQMのエラー・カテゴリーとを2015年に統合したもの。

上位レベルのエラー・カテゴリーはMQMと同じだが、細かな部分で相違もある。たとえば重大度にはDQF独自の「Kudos」（称賛）が入っている。Kudosでは翻訳ミスで「減点」するのではなく、期待以上の出来栄えに「加点」をするのに使う。

ITS（Internationalization Tag Set）バージョン2.0

ウェブ技術の標準化団体であるW3Cによる標準。XLIFFなどで翻訳品質上の問題をマークするときに使うエラー・カテゴリーが設けられており、MQMを参考にしている。

ISO 21999

国際標準化機構による規格。翻訳品質評価のモデルおよびメトリクスを提示している。2018年4月時点では策定中である。

ASTM WK46396およびWK54884

国際標準化団体であるASTM Internationalによる翻訳品質評価の手法およびメトリクス。2018年4月時点でともに策定途中である。

Garvinに基づく翻訳品質5分類

Fieldsら（2014）は論文で、経営学者Garvin（1984）の論に基づき、翻訳の「品質」を5つに分類した。超越的、プロダクトベース、ユーザーベース、生産ベース、価値ベースである。西野（2016a）による解説を以下に簡単に紹介するとともに、5つの関係図を示す。

- 超越的：専門家が主観的に評価。
- プロダクトベース：客観的な指標で評価。
- ユーザーベース：ユーザーが主観的に評価。
- 生産ベース：あらかじめ定めた要件や仕様をどの程度満たすかで評価。
- 価値ベース：費用と便益の比較で評価。

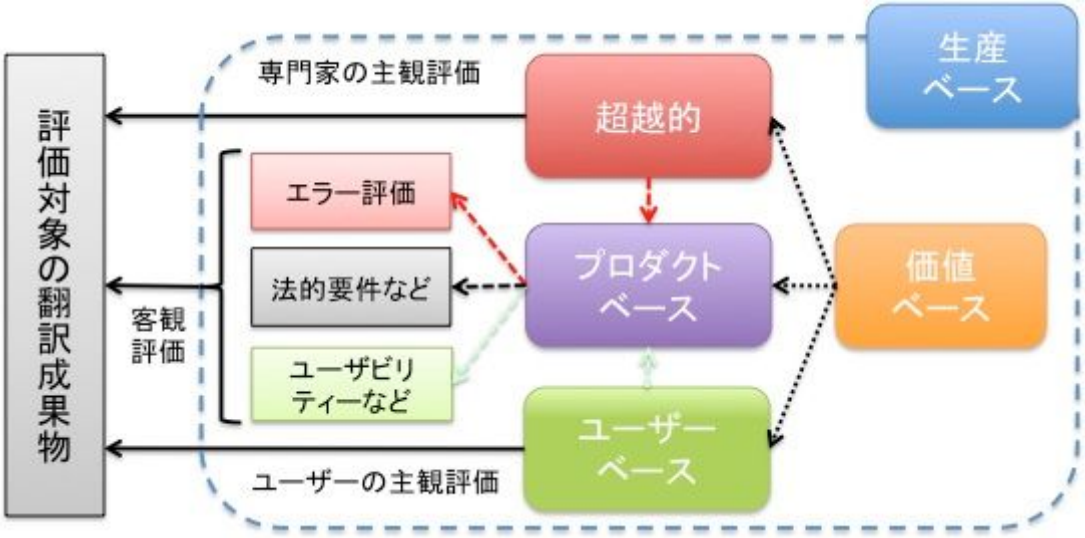

図2：Garvinに基づく翻訳品質評価アプローチ （西野2018より）

図2の中で、左端の翻訳成果物を直接的に評価しているのは「超越的」、「プロダクトベース」、「ユーザーベース」の3つである。価値ベースは費用による評価の相対化である。たとえば、翻訳自体はいくら良くても、コストが膨大にかかっていれば相対的な評価は下がる。

また生産ベースは「あらかじめ定めた要件や仕様をどの程度満たすか」で評価するため、直接的には翻訳成果物を評価できない。直接的に評価するためには、どのエラー評価手法（プロダクトベース）を使うか、コストはどの程度か（価値ベース）といった点を関係者間で合意する必要がある。ほかの4つよりも上位に位置するため、点線で囲っている。

この分類で言うと、本ガイドラインの「JTF翻訳品質評価モデル」は、仕様を重視する生産ベースの考え方を基本にしつつ、具体的な評価手法としてプロダクトベースに属するエラー評価の仕組みも備えているということになる。

参考文献

- ASTM International, "ASTM WK46396 New Practice for Development of Translation Quality Metrics," https://www.astm.org/DATABASE.CART/WORKITEMS/WK46396.htm (アクセス：2018-09-30).
- ASTM International, "ASTM WK54884 New Guide for Public Language Quality Assessment (LQA) Methodology," https://www.astm.org/DATABASE.CART/WORKITEMS/WK54884.htm (アクセス：2018-09-30).
- P. Fields, D. Hague, G. S. Koby, and A. Melby, “What Is Quality? A Management Discipline and the Translation Industry Get Acquainted,” Rev. Tradumàtica, no. 12, pp. 404–412, 2014.
- D. Garvin, "What Does “Product Quality” Really Mean?," Sloan Management Review, vol: 26 (1) pp: 25-43, 1984.
- ISO, "ISO/AWI 21999 Translation quality assurance and assessment -- Models and metrics", https://www.iso.org/standard/72345.html (アクセス：2018-09-30).
- A. Lommel, A. Burchardt, and H. Uszkoreit, Eds., “Multidimensional Quality Metrics (MQM) Definition v1.0,” 2015.
- SAE International, "Translation Quality Metric," J2450_201608, 2016.
- TAUS, "DQF & QT21", https://www.taus.net/evaluate/qt21-project (アクセス：2018-09-30).
- W3C, "Internationalization Tag Set (ITS) Version 2.0," https://www.w3.org/TR/its20/ (アクセス：2018-09-30).
- 西野竜太郎, “翻訳品質のランチボックス：翻訳の「品質」とは（2）,” 日本翻訳ジャーナル, no. 284, 日本翻訳連盟, pp. 22–23, 2016a.
- 西野竜太郎, “翻訳品質のランチボックス：従来のエラーベース評価手法,” 日本翻訳ジャーナル, vol. 285, pp. 26–27, 2016b.
- 西野竜太郎, “翻訳品質のランチボックス：仕様とJTF翻訳品質評価ガイドライン,” 日本翻訳ジャーナル, vol. 294, pp. 24–25, 2018.

関連情報

改訂履歴

- 第1版発行（2018-11-15）

日本翻訳連盟について

一般社団法人日本翻訳連盟（JTF）は、翻訳に関わる企業、団体、個人の会員からなる産業翻訳の業界団体です。

ウェブサイト：http://www.jtf.jp/

制作者について

JTF翻訳品質委員会の部会である「翻訳品質評価ガイドライン検討会」が制作しました。同検討会は当時JTF専務理事だった故・田中千鶴香氏が2016年度に招集し、2018年度に第1版を完成させました。各年度の委員は以下の通り（五十音順）です。

2018年度

- 東尚子（個人翻訳者）
- 工藤竜広（株式会社サン・フレア）
- 久保田真司（パナソニックIPマネジメント株式会社）
- 高橋聡（JTF専務理事）
- 田嶌奈々（株式会社翻訳センター）
- 外山厚子（株式会社ホンヤク社）
- 西野竜太郎（JTF理事）※ 委員長およびガイドライン編集責任者
- 舟津由美子（個人翻訳者）
- 三笠綱郎（株式会社十印）
- 吉川未希子（株式会社ヒューマンサイエンス）
- 若林啓子（株式会社クロスランゲージ）

2017年度

- 東尚子（個人翻訳者）
- 工藤竜広（株式会社サン・フレア）
- 久保田真司（パナソニックIPマネジメント株式会社）
- 高橋聡（JTF専務理事）

- 田嶌奈々（株式会社翻訳センター）
- 田中千鶴香（JTF専務理事）※ 委員長
- 外山厚子（株式会社ホンヤク社）
- 西野竜太郎（JTF理事）※ 委員長（2017年9月～）およびガイドライン編集責任者
- 舟津由美子（個人翻訳者）
- 松下慶子（株式会社クロスランゲージ）
- 三笠綱郎（株式会社十印）
- 吉川未希子（株式会社ヒューマンサイエンス）
- 若林啓子（株式会社クロスランゲージ）

2016年度

- 東尚子（個人翻訳者）
- 河野弘毅（ポストエディット東京）
- 高野敬一（株式会社ヒューマンサイエンス）
- 高橋聡（JTF理事）
- 田嶌奈々（株式会社翻訳センター）
- 建部優子（株式会社十印）
- 田中千鶴香（JTF専務理事）※ 委員長
- 辻元正明（株式会社クロスランゲージ）
- 西野竜太郎（合同会社グローバリゼーションデザイン研究所）
- 吉川未希子（株式会社ヒューマンサイエンス）

著作権

本ガイドラインの著作権は、一般社団法人日本翻訳連盟（JTF）が保有しています。
Copyright ©2018 Japan Translation Federation

利用許諾

本ガイドラインは、クリエイティブ・コモンズ・ライセンス「表示 4.0 国際（CC BY 4.0）」の下で利用が可能です。
同ライセンスの要約：https://creativecommons.org/licenses/by/4.0/deed.ja

免責

本ガイドラインおよび関連文書は現状有姿で提供され、日本翻訳連盟は、その正確性や使用について一切の保証をいたしません。これらの文書には、誤りや不正確な情報が含まれる

可能性があります。日本翻訳連盟は、事前の通知なくこれらの文書の記載内容を変更する権利を有します。

フィードバック

本ガイドラインに関するご意見などがあれば日本翻訳連盟までお送りください。
メールアドレス：info@jtf.jp

クレジット

本ガイドラインの一部エラー項目名は、CC BY 4.0ライセンス（https://creativecommons.org/licenses/by/4.0/）に基づき、DFKIおよびQTLaunchPadによる「Multidimensional Quality Metrics (MQM) Definition」（http://www.qt21.eu/mqm-definition/）を翻訳して利用しています。

Some error category names in this guideline are translated from "Multidimensional Quality Metrics (MQM) Definition" (http://www.qt21.eu/mqm-definition/) by DFKI and QTLaunchPad, based on CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Copyright ©2015, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH/German Research Center for Artificial Intelligence (DFKI) and QTLaunchPad.