

実務翻訳における翻訳品質評価プロセス

田中千鶴香

日本翻訳連盟専務理事/個人翻訳者

本稿では、実務翻訳で行われる翻訳品質評価プロセスについて、関連する用語やワークフローを説明する。2016年に日本翻訳連盟 (JTF) が実施した「翻訳品質評価方法に関する業界アンケート」¹の結果を参照しながら、特に日本の実務翻訳市場における翻訳品質評価プロセスの実状を紹介する。

以下で言及する作業はすべて人手によるもので、機械による自動評価等を含まない。原則として、原文と訳文を照合して品質を評価するバイリンガルチェック作業について述べている。

また、翻訳作業を担当する人的リソースの能力審査にも言及するが、それらは、翻訳を担当する翻訳者や翻訳会社の能力評価が目的であって、人材の教育を目的としていない。

1 翻訳品質評価プロセスに関連する用語

翻訳品質評価プロセスに関連する用語として、品質保証 (Quality Assurance)、品質評価 (Quality Assessment/Evaluation)、品質検査 (Quality Check) を挙げることができる。ただし、どの用語にも明確な定義がなく、社内や部署内などの限られた関係者間の合意に基づいて、それぞれの用語が使用されているのが現状である。

JTF が実施した「翻訳品質評価方法に関する業界アンケート」の結果をみると、「翻訳の品質を評価する」という作業について、アンケートの回答者が実際にイメージした作業内容は一様ではない²。用語の定義と同様に、翻訳品質評価に関する具体的な作業内容についても、業界共通の認識がないと思われる。後述のように、この状況は海外でも同じである。

1.1 品質保証 (Quality Assurance)

最もよく目にする用語は「品質保証 (Quality Assurance)」である。米国の言語サービス業界誌 MultiLingual のグロッサリーでは以下のように定義されている。

Quality Assurance (QA): the activity of providing evidence needed to establish confidence among all concerned that quality related activities are being performed effectively. All those planned or systematic actions necessary to provide adequate confidence that a product or service will satisfy given requirements for quality³.

欧州委員会の支援を受ける研究者グループ QTLaunchPad⁴ が作成した Multidimensional Quality Metrics (MQM)⁵ では、以下のように定義されている。

Quality assurance refers to ways of preventing mistakes or defects in manufactured products and avoiding problems when delivering solutions or services to customers. Quality assurance relies on continual assessment of quality⁶.

翻訳会社がサービスプロバイダーとして機能する機会の多い市場 (たとえば日本) の場合は、「品質保証」が「内部/外部委託により作成した成果物の品質を発注元に納める前に保証するプロセス」を指すことが多い。具体的には、翻訳ワークフローの内部で、その最終段階に行われる、納品前の詳細な確認を行う工程であり、上記の MQM の定義に近い。ただし、成果物の問題点 (エラー) の検出と排除は行われるが、品質の定量的評価はあまり行われない。

MQM には、Quality Assurance と Quality Assessment をどのように区別して運用するかについての言及はない。ISO/NP21997⁷ の準備資料 (非公開) においても、それぞれの意味が不明確なまま Quality Assurance と Quality Assessment が使用されている。今後 ISO NP 21999 に関する議論が進む過程で、各用語が明確に定義されることが期待される。

¹日本翻訳連盟「翻訳品質評価方法に関する業界アンケート 結果報告書」http://www.jtf.jp/jp/style_guide/pdf/2016JTF_TranslationQualitySurvey.pdf

²同上の設問 11

³MultiLingual, 2016 RESOURCE DIRECTORY / INDEX

⁴<http://www.qt21.eu/>

⁶<http://www.qt21.eu/mqm-definition/definition-2015-06-16.html#quality-terms>

⁷ISO/NP2199 Translation quality assurance and assessment - Models and metrics

1.2 品質評価 (Quality Assessment/Evaluation)

翻訳成果物の品質を評価する工程を指す最も一般的な用語で、訳文の問題点に対するスコアリング (点数化) を伴う定量的な評価を意味する場合がある。ただし翻訳品質評価において常に定量評価がなされるわけではない。MQM の定義も定量評価には言及していない。

Quality assessment (sometimes considered synonymous with “quality evaluation”) is the measurement of the extent to which a product complies with quality specifications.

品質評価は、翻訳ワークフローの内部と外部のいずれでも行われる。前者のケース (ワークフロー内で実施) では、品質保証と混同されるケースが多い。後者の場合 (ワークフロー外で実施) は、翻訳を担当したリソースの能力審査を目的とする場合が多い。

1.3 品質チェック (Quality Check)

チェックフォームなどを用いた、手動による翻訳品質の確認作業を指すことが多い。国内では比較的広く行われている翻訳品質の確認作業である。定量評価を伴う場合もある。品質保証、品質評価との作業区分があいまいで、各組織が独自の判断で作業内容を規定している場合が多い。

2 翻訳の品質を表す用語

翻訳品質の定量評価で用いられる、翻訳の品質を表す用語を説明する。評価方法が成果物の問題点に着目する減点方式 (エラーベース) であるため、問題点の種類 (type/category) を表す用語 (正確さ、流暢さなど) と問題点の重大度 (Severity) を表す用語 (軽度、重度など) が用いられる。

問題点の種類を表す用語は、評価項目として用いられる。評価項目は、複数の階層に分けられる場合がある。

2.1 問題点の種類 (type/category) を表す用語

翻訳業界には、業界標準と言われる評価モデルがあるが、あまり浸透しておらず⁸、大半の企業が独自に作成した評価モデルを使用し、それぞれに評価項目を設けているため、さまざまな評価項目が用いられている。ただし、以下の「正確さ (Accuracy)」、「流暢さ (Fluency)」、「用語 (Terminology)」、「スタイル (Style)」の4項目は、ほぼすべての評価モデルで採用されている。

2.1.1 正確さ (Accuracy)

正確さ (Accuracy) は、原文の情報が訳文に正しく反映されているかどうかを評価する項目である。謝った訳になっている、原文にある情報が存在しない、原文にない情報が存在するといった問題点がこの項目に分類される。

正確さは、主に成果物である訳文の適切性によって判断される。たとえば、原文と訳文の意味内容が表面的には離れていても、情報が適切に反映されていれば、問題点とみなされない。

2.1.2 流暢さ (Fluency)

流暢さ (Fluency) は、一般に、訳文言語を母語とするネイティブ話者が訳文を読んだときに、スムーズに読めるかどうかを評価する項目である。ただしその定義はあいまいで、正確に訳されているが読み物としての読みごたえのない状態をこの項目に分類する考え方もあれば、文法や句読法の間違いをこの項目に分類する考え方もある。

後述する TAUS DQF⁹ と MQM の評価項目統合のような活動があることから、流暢さ (Fluency) についてもその定義や概念が徐々に統一されることが期待される。

2.1.3 用語 (Terminology)

特定の専門分野で用いられるべき用語が訳文で適切に使用されているかどうかを評価する項目である。

2.1.4 スタイル (Style)

テキストの文体や表記が適切であるかどうかを評価する項目である。たとえば日本語の場合は漢字かな表記の不統一などが該当する。個人の文章表現上の特色といった客観的に判断できないものは評価の対象としない。

⁸日本翻訳連盟「翻訳品質評価方法に関する 業界アンケート 結果報告書」http://www.jtjf.jp/jp/style_guide/pdf/2016JTf_TraslationQualitySurvey.pdf の設問 9

⁹<https://www.taus.net/evaluate/dqf-tools>

2.2 問題点の重大度 (Severity) を表す用語

問題点の影響の大きさに応じて重大度 (Severity) を判定する作業が行われる場合がある。この判定には、通常、軽度 (Minor)、重度 (Major)、深刻 (Critical) などのような3~4段階のラベルが用いられる。どのラベルを付けるかは、その問題点が発注元や最終読者にもたらす影響 (安全・法律・金銭的な不利益、誤読や誤理解の発生) に応じて判断され、原文と訳文との齟齬の大きさはほとんど考慮されない。つまり、原文と意味内容がかけ離れた訳文が存在しても、周囲に与える影響が軽微であれば、重大度は低いと判断される。

2.3 用語に関する業界の取り組み

業界標準の評価モデルについては、現在、モデル間で異なっている評価項目を対応付けて、統合する試みが進められている。TAUS DQF で使用するメトリックスと MQM のメトリックスの統合はかなり進んでおり¹⁰、それらと SAE J2450 (米国の Society of Automotive Engineers が自動車業界向けに策定した翻訳品質評価モデル)¹¹ とのマッピングも行われている。

TAUS DQF と MQM の評価項目を統合した最新のモデルが TAUS のウェブサイトで公開され¹²、今後も更新される予定である。

3 業界標準の評価モデル

業界標準の評価モデルには以下がある。

- LISA の QA モデル¹³
- SAE J2450
- TAUS DQF

企業が独自に作成して使用している評価モデルの大半は、上記の LISA QA モデルや SAE J2450 をベースにしたものである。

4 翻訳品質評価の目的

翻訳の品質評価は、成果物の納品前後の検品のためだけでなく、翻訳会社、翻訳者、チェック担当者といったリソースの能力審査のためにも行われる。多くの評価モデルがどちらの目的にも使えるように設計されている。

前述の SAE J2450 は、翻訳成果物を純粋な「プロダクト」とみなして評価するモデルである。そのため人的リソースの能力審査のためには使用されない。

5 翻訳品質評価の一般的なワークフロー

翻訳品質を定量評価する場合のワークフローを以下に説明する。評価作業を担当する人的リソースについても触れる。

5.1 事前準備

評価対象の特定と評価担当者の選定が行われる。必要なファイル類には以下がある。

- 評価フォーム (Excel 形式など)
- 評価担当者向けの指示書
- 翻訳担当者が翻訳時に参照した指示書、用語集、スタイルガイド

評価担当者向けの指示書には、「評価手順」、「評価基準」、「評価時の注意点」などが書かれている。

¹⁰QT21 Deliverable 3.1 Harmonised Metric <http://www.qt21.eu/wp-content/uploads/2015/11/QT21-D3-1.pdf>

¹¹<http://www.sae.org/standardsdev/j2450p1.htm>

¹²<https://www.taus.net/evaluate/qt21-project#harmonized-error-typology>

¹³LISA QA Model (Localization Industry Standards Association Quality Assurance Model)。LISA は国際的なローカリゼーション業界団体だったが現在は活動していない。

5.2 評価の実施

担当者が評価フォームを使用して、翻訳の品質を評価する場合の一般的な手順は以下のとおりである。問題点の発見、分類、および重大度の判定を行う、いわゆる「減点方式」による評価である。多くの場合、自由記述された評価担当者自身の見解（コメント）も評価結果の重要な一部となる。

- (1) 評価担当者が原文と訳文を照合して訳文の問題点を発見し、当該箇所の原文と訳文を評価フォームの所定の欄に記入する。
- (2) 発見した問題点を評価項目のいずれかに分類し、重大度を判定する。
- (3) 訳文に誤りがある場合は、正しい（あるいは改善した）訳文を評価フォームの所定の欄に記入する。
- (4) 必要に応じて、判断の根拠を評価フォームの所定の欄に記述する。
- (5) 成果物全体の品質に対するコメントを評価フォームの所定の欄に記述する。

評価の判断基準は、評価担当者向けの指示書に詳細に書かれている。主観的評価を避けるための指示も具体的に書かれていることが多い。

定量評価を行う場合は、問題点が分類され、重大度が評価フォームに記入されると、問題点の総数と個々の評価項目と重大度に付与された重み（Weight）に応じた「点数」が計算される。翻訳を担当したリソースに結果をフィードバックするなどの用途を考慮すると、数値による評価結果を得るだけでは不十分で、評価担当者の自由記述によるコメントも得る必要がある。

場合によって、問題点としてはカウントしないが関係者が留意すべき事象が評価対象の訳文（または原文）に見つかり、それらを所定の評価項目に分類して、評価フォームに記入する。「恣意（Preferential）」、「中立（Neutral）」といった評価項目が用いられることが多い。

5.3 結果の活用

評価結果は、将来の品質の改善やリソースの能力審査のために活用される。

5.4 評価を担当する人的リソース

翻訳品質の評価を担当するのは、発注企業や翻訳会社の社員が多い。さらに外部のリソースに業務委託する場合もある。一般に、経験と実績のある翻訳者またはレビュー（チェック）担当者が評価を担当する。

6 まとめ

日本の翻訳業界で行われている品質評価手法を概観した結果を以下のようにまとめる。

- 業界標準の評価モデルがあまり浸透しておらず、企業独自の評価モデルが使用されている。
- 翻訳品質評価で用いられている用語の定義が明確でない。そのため、同じ用語で表される作業（たとえば「品質保証」）であっても、実際の作業内容が同じとは限らない。
- 「正確さ（Accuracy）」、「流暢さ（Fluency）」、「用語（Terminology）」、「スタイル（Style）」の4つの評価項目は、ほぼすべての評価モデルで採用されている。
- 海外の業界標準評価モデルでは、評価項目の整理と統合が進められている。
- 翻訳の品質評価は、成果物の納品前後の検品だけでなく、リソースの能力審査のためにも行われる。多くの評価モデルがどちらの目的にも使えるように設計されている。
- 大半の評価方法は、成果物の問題点に着目する減点方式（エラーベース）である