

アノテーションに使用した誤りの分類体系

藤田 篤
情報通信研究機構

2024 年 10 月 17 日

所与の誤りを、次に示す 18 種類のカテゴリのいずれかに分類する。リストの上から順に各カテゴリに該当するか否かを判定することで、複数のカテゴリに該当すると判断されるような場合でも、最初に該当すると判断されたカテゴリ 1 つのみを付与する。なお、特筆する場合を除き、MT 訳中の該当箇所のみをアノテートする（原文中の対応箇所のアノテートは不要）。

Terminology（用語）

特定の専門分野で用いられるべき用語、人名・組織名などの固有表現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。Accuracy（正確さ）や Fluency（流暢さ）よりも優先的に該当するか否かを判断する。

[1] Terminology/Inappropriate for content

用語の訳として文脈に適した表現が使用されていない。

- (1) 原文: This mainly happens in the small intestine.
MT 訳: これは主に小さな腸で行われる。
判断理由: いかなる文脈でも小腸のことを「小さな腸」とは言わない。
- (2) 原文: Foreign Ministry spokesman Noh Kwang-II said, "..."
MT 訳: 外務省広報官のNoh Kwang-II氏は次のように述べた。「...」
判断理由: 未翻訳 (Accuracy/Untranslated text) とも取れるが、こちらを優先する。
- (3) 原文: They are struggling to deal with a spike in cybertheft attacks on the nation's online banking system.
MT 訳: 彼らはインターネットバンキングシステムへのサイバー攻撃に対処しようと必死だ。
判断理由: “theft” の欠落 (Accuracy/Omission) とも取れるが、こちらを優先する。

[2] Terminology/Inconsistent use

用語の訳が一貫していない。

- (4) 原文: Delta introduces ... Delta is upping ...
MT 訳: デルタ航空は ... デルタは ...
判断理由: 略称や別表記を特に導入することなく、異なる訳出をしている。

Accuracy（正確さ）

Accuracy（正確さ）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。誤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をこの項目に分類する。例えば、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、問題点とみなさない。

[3] Accuracy/Untranslated text

原文中の翻訳すべき情報が翻訳されないまま MT 訳に含まれている。

- (5) 原文: The use of new unmanned drones could drastically change the nature of the fight.
MT 訳: 新型の無人ドローンは、戦いの本質をdrasticに変えうる。

[4] Accuracy/Addition

原文中にない情報が MT 訳で付加されている場合に該当する。

- (6) 原文: Once food is liquid, it can pass into the blood.
MT 訳: 食べ物はひとたび液状になると、血液中を簡単に流れることができる。
判断理由: 原文では様態は示されていない。
- (7) 原文: These engineers can work in a broad range of settings and enjoy lucrative pay.
MT 訳: これらの技術者は様々な状況下で働くことができるため高額な月給をもらっている。
判断理由: 給与の支払い頻度については言及されていない。

[5] Accuracy/Omission

原文中の情報が MT 訳中で欠けている場合に該当する。原文中の対応箇所をアノテートする。ただし、自明な要素の省略など、その情報の有無によらず内容に変化がない場合は、この誤りには含めない。

- (8) 原文: There are possibly as many as 20.7 million cases — including many children.
MT 訳: 最大で 2070 万件にも上る可能性があり、その中には子どもの情報も含まれている。
判断理由: “many” が訳出されていないことが問題である。
- (9) 原文: They are struggling to deal with a spike in cybertheft attacks on the nation's online banking system.
MT 訳: 彼らはインターネットバンキングシステムへのサイバー窃盗に対処しようと必死だ。
判断理由: “nation's” が訳出されていないことが問題である。

[6] Accuracy/Mistranslation

原文中の情報が誤って訳されている場合に該当する。

- (10) 原文: Once food is liquid, it can pass into the blood.
MT 訳: かつて食べ物は液状であり、血液中を流れることができる。
判断理由: “Once” の誤解。
- (11) 原文: I can arrange to protect my children most of the time.
MT 訳: ほとんどの時間、子供を守ることができる。
判断理由: “time” の誤解。

- (12) 原文: The special rule is applied to this case.
MT 訳: この事例には特別なルールが適応される。
判断理由: 誤変換 (Fluency/Spelling) ではなく語選択の誤りなので致命的である。
- (13) 原文: "...," the firm said.
MT 訳: 「...」と同社は述べる。
判断理由: 時制の誤り (これは文法ではなく内容の解釈の誤りである)。
- (14) 原文: The renewal of power in France was soon followed with the invasion of Italy.
MT 訳: イタリア侵攻の直後フランスの国力は再生した。
判断理由: 事象の生起順序の誤解。
- (15) 原文: The sites are disappearing because the cost of making the structures strong enough to withstand earthquakes is too high.
MT 訳: 耐震工事の費用がかかるため、その計画も消えかかっている。
判断理由: 節の誤解。
- (16) 原文: The move, expected for several weeks, drew swift opposition from China.
MT 訳: この動きは数週間に渡ると予想されていたが、即座に中国から反対の声が挙がった。
判断理由: 節の誤解。

Fluency (流暢さ)

Fluency (流暢さ) は、目標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。文法や句読法の間違ひのみでなく、内容は正確に伝わるが難解である状態や目的に照らして適切ではない場合もこの項目に分類する。

[7] Fluency/Punctuation

句読点を含む約物の問題で流暢さが損なわれている。

- (17) 原文: The change is welcome news to the United States, which maintains troops in Japan.
MT 訳: この変化は日本に軍を置く、米国にとってはうれしい知らせである。
判断理由: 不適切な位置の読点により解釈が困難になっている。
- (18) 原文: "At the moment, they're not going to be launched," said AAA.
MT 訳: 「今のところ始動する予定はありません。」と AAA は述べた。
判断理由: 翻訳の仕様によらず一般に直接引用の末尾に句点を用いないため、Style/Awkward ではなくこちらに分類する。

[8] Fluency/Spelling

綴りの誤りがある。かな漢字変換誤りなど、発音の類似性に関連するもののみを含む。

- (19) 原文: He regrets the reality that he had no chance to meet me earlier.
MT 訳: もっと早くに会う機械がなかった。
- (20) 原文: Avoid wearing black shoes with a light suit.
MT 訳: 明るい色ののスーツの時は黒い靴は避ける。

[9] Fluency/Grammar

語の活用、助詞の選択などの文法的な問題がある。

- (21) 原文: When I went to the store, they had a new watch.
MT 訳: その店に行ったら、新しい時計が売っていた。
判断理由: いわゆる「ら抜き」も活用の誤りとみなす。
- (22) 原文: Those who cannot change their minds cannot change anything.
MT 訳: そして、考え方を変えれない人は、何も変えることができない。
判断理由: いわゆる「ら抜き」も活用の誤りとみなす。
- (23) 原文: The opportunities to study English and have contact with people from other countries are fairly numerous.
MT 訳: 英語を学んだり、外国の人々と交流する機会はかなり多い。
判断理由: 「～たり～たり」という呼応表現を適切に使うことができていない。
- (24) 原文: They duck their heads into the water, trying to catch the apple with their teeth.
MT 訳: 彼らは頭を水に突っこみ、りんごを歯でつかもうとする。
判断理由: コロケーションが不自然である。

[10] Fluency/Cohesiveness

節間、文間の結束性が損なわれている場合に該当する。自明な要素の有無により内容に変化がない場合でも、訳出されていないために結束性が損なわれている場合は、原文中の対応箇所をアノテートする。

- (25) 原文: Once food is liquid, it can pass into the blood.
MT 訳: 食べ物はひとたび液状になると、それは血液中を流れることができる。
判断理由: 日本語では自明な主語を省略するのが一般的である。
- (26) 原文: The group has already unveiled new weapons, ...
MT 訳: そのグループはすでに新型兵器を投入し、 ...
判断理由: “The group” が何を指すかを前方文脈から補完する必要がある（という理由で誤りと認定した場合の分類）。

[11] Fluency/Register

分野・文書タイプに適した語句が使われていない。

- (27) 原文: Food is digested. This means that it is changed into a simpler, liquid form.
MT 訳: 食べ物は消化されます。てことは、食べ物はただの液体状に変えられるということです。
判断理由: ぐだけすぎている。
- (28) 原文: ..., sources said Thursday.
MT 訳: ... が木曜日に発覚した。
判断理由: 日本語の報道記事などでは、曜日ではなく日付を記述する。

[12] Fluency/Inconsistency

（用語以外の）表現について、訳出（語選択）が一貫していない。同一の語の表記の揺れは後述の Style/Awkward に含める。

- (29) 原文: The proposed method processes ... The method consists of ...
MT 訳: 提案手法は、... この方法は ...
判断理由: 同じ指示対象を「手法」と「方法」というように異なる表現で訳出している。

[13] Fluency/Others

上記以外の理由で流暢さが損なわれている。

Style (スタイル)

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。個人の文章表現上の選好など、客観的に判断できないものは評価の対象としない。

[14] Style/Awkward

文法や語句の選択には問題ないが、文字や表記などいわゆるスタイルで扱う要素が適切ではない。

- (30) 原文: AAA purchased the data sometime from late April to May from another list trader.
MT 訳: AAA は 4 月下旬～5 月に別の名簿業者から情報を購入した。
判断理由: 報道記事ではこのような記号を使用しない。
- (31) 原文: The cabinet has authorised the army to mobilise 40,000 reservists.
MT 訳: 同国の内閣は軍に40000 人もの予備兵を召集する権限を与えた。
判断理由: 「40000 人」は文法的だが、報道記事においては数字の表記に決まりがある。
- (32) 原文: Fig. 2 shows ... Fig. 3 demonstrates that ...
MT 訳: 図 2 は... を表す。図 3 は... を表わす。
判断理由: 送り仮名が統一されていない。
- (33) 原文: For example, ... For example, ...
MT 訳: 例えば...。たとえば...。
判断理由: 漢字表記とかな表記が統一されていない。
- (34) 原文: Reports were given on suspected cases of child abuse. We are currently looking for countermeasures.
MT 訳: 虐待があるという報告があった。現在改善策を検討しています。
判断理由: 敬体と常体が混在している。

Locale (ロケール)

目標言語が使用される地域における慣例に適合しているかどうかを評価する項目である。

[15] Locale/Format

人や組織などの名前や敬称、住所、通貨、日付、時間、度量衡、数字、電話番号などの表記に問題がある。

- (35) 原文: In the first year of MLB, Ohtani signed the league-minimum salary of \$545,000.
MT 訳: 大谷の MLB 最初の年の年俸はリーグ最低保証額の545,000 ドルだった。
判断理由: 日本語では米ドル表記よりも日本円換算額を示す方が良い。

[16] Source error

原文中に適切な翻訳が不可能な表現がある。

[17] Non-translation

文全体が原文と無関係であるまたは個々の誤りを識別できない。

[18] Other errors

上記以外の問題である。