

MQM-like issue typology used for annotation

Atsushi Fujita

Advanced Translation Technology Lab., ASTREC, NICT

October 17, 2024

In this task, the worker is asked to classify each issue into one of the 18 categories listed below. By following the decision list, a single category is determined for each issue, even if the issue fulfills conditions of multiple categories. Unless otherwise noted, the issues are annotated in the translation side; their corresponding parts in the source side do not need to be annotated.

Terminology

This top category evaluates translations of terms in a specific sub-language and named entities, such as person names and organization names. This is prioritized over accuracy and fluency issues in translation.

[1] Terminology/Inappropriate for content

This category is used if the translation of a term or an named entity is not properly translated.

- (1) Source: 平成 13 年に公開された千と千尋の神隠しは宮崎駿の代表作だ。
MT output: Sen to Chihiro no kamikakushi released in 2001 is a representative work of Shun Miyazaki.
Explanation: The movie title in English is “Spirited Away” and the first name of the director is “Hayao.” Note that the entire person name should be annotated for the second issue.
- (2) Source: これは東京大学との共同プロジェクトの成果だ。
MT output: This is the outcome of the joint project with Tokyo University.
Explanation: Proper translation of the organization name is “The University of Tokyo.”
- (3) Source: 永住許可を得ると、無期限に滞在できます。
MT output: Once you obtain a permission for permanent stay, you can stay indefinitely.
Explanation: Proper translation of the term is “Permission for Permanent Residence.”

[2] Terminology/Inconsistent use

The same term/entity is not consistently translated. Some mentions would be translated with pronouns (e.g., “it”) or hypernyms (“the company”) but they should not be judged erroneous as long as they refer to the proper entity.

- (4) Source: デルタ航空は ... デルタ航空はまた ...
MT output: Delta Air Lines introduces ... Delta is also starting ...
Explanation: The same entity is translated into different ways without a proper introduction of the short name.

Accuracy

This top category evaluates if the translation is accurate with respect to the information that the source conveys. Even if the translation is dissimilar from the source regarding structure and level of figurativeness, it should be judged accurate as long as it conveys the same meaning as the source.

[3] Accuracy/Untranslated text

An expression in the source remains in the target text.

(5) Source: 初対面の人に対しては一般的に「はじめまして」と挨拶します。

MT output: When we meet a person first time, we typically say “はじめまして.”

[4] Accuracy/Addition

Content not present in the source is wrongly added in the translation.

(6) Source: 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。

MT output: There are possibly as many as 20.7 million cases, including many children.

Explanation: The quantity is not mentioned in the source.

(7) Source: 改修により講堂の耐震性が増した。

MT output: Through the repair, the earthquake resistance of the auditorium is substantially improved.

Explanation: The translation wrongly adds the degree of improvement.

[5] Accuracy/Omission

Content present in the source is wrongly omitted in the translation. An omission is not an issue if the meaning of the source is properly conveyed. Note that issues of this type are annotated at the source side.

(8) Source: 食べ物は一たび液状になると、血液中を簡単に流れることができる。

MT output: Once food is liquid, it can pass into the blood.

Explanation: The adverbial information is omitted in the translation.

(9) Source: 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。

MT output: These engineers can work in a broad range of settings and enjoy lucrative pay.

Explanation: The frequency of the payment is not presented in the translation.

[6] Accuracy/Mistranslation

Content present in the source is misrepresented in the translation.

(10) Source: 銀行は昨日行った。

MT output: The bank went yesterday.

Explanation: Misunderstanding of the omitted subject.

(11) Source: 改修により講堂の耐震性が増した。

MT output: Through repairs, the earthquake resistance of the auditorium is improved.

Explanation: Single time of repair was mistakenly translated into plural repairs.

- (12) Source: その失敗をした瞬間、新しいアイデアが閃いた。このとき彼は「...」と述べる。
 MT output: When he made that mistake, a good idea leaped into his mind. At this moment, "..." he say.
 Explanation: Wrong tense (the basic form, but the Japanese verb indicates past tense in this context). This issue is not carried over until Fluency/Grammar since it stems from misunderstanding of the content.
- (13) Source: 株式会社 XXX の AAA 先生を講師としてお招きしました。
 MT output: We invited Professor AAA of XXX Co., Ltd. as a lecturer.
 Explanation: “先生” in the source is a honorific expression and does not mean the title at universities and colleges.
- (14) Source: イタリア侵攻の直後フランスの国力は再生した。
 MT output: The renewal of power in France was soon followed with the invasion of Italy.
 Explanation: Misunderstanding of the relationship between two propositions.

Fluency

This top category evaluates if translations are fluent enough. It includes not only grammatical issues but also readability issues.

[7] Fluency/Punctuation

Wrong punctuation makes the sentence influent.

- (15) Source: 店長は、「本件に関し、反省しております」と謝罪した。
 MT output: The store manager apologized, “This will never happen again”.
 Explanation: This fullstop must be inside the quote in American English. This is not a Style/Awkward issue because it does not depend on the specification of translation.
- (16) Source: 大腸がんのリスクとして、肥満、喫煙、運動不足などがあります。
 MT output: There are some of the risk factors for colorectal cancer; obesity, smoking, and a lack of exercise.
 Explanation: Incorrect use of a semi-colon.

[8] Fluency/Spelling

Misspelling including those based on similar pronunciation, missing/wrong hyphenation, case errors (for other than terms), and diacritic errors.

- (17) Source: 信頼性が低下している可能性があります。
 MT output: Its possible that it's reliability is degraded.
- (18) Source: 多言語学習では言語間の共通点と相違点を分析することが重要です。
 MT output: In multi-lingual learning, it is important to analyze the similarities and differences between languages.
 Explanation: No hyphen is needed.
- (19) Source: 詳細はリンク先のウェブサイトをご覧ください。
 MT output: For details, please refer to the linked web site.
 Explanation: No whitespace is needed.

[9] Fluency/Grammar

Issues that hurt grammaticality, such as inflection of verbs, agreement, misuse of prepositions

- (20) Source: 夏の間、この地域にある公立学校のプールは一般市民に開放されます。
MT output: During the summer, the municipal school pools in this area are open to public.
Explanation: In this context this part should be “the public.”
- (21) Source: 我々は、このプロジェクトがどのように社会に貢献してきたかについて話し合った。
MT output: We discussed how this project has contributed to the society.
Explanation: In this context, no determiner is necessary.
- (22) Source: 医者がいつもどおりの時間に来ていたら、患者は助かったかもしれない。
MT output: The patient might have survived if the doctor came at the usual time.
Explanation: Verb tense in the subjunctive clause is wrong. This issue is not classified as Accuracy/Mistranslation.
- (23) Source: 多くの市民がその法案に反対した。
MT output: Many citizens opposed to the bill.
Explanation: “oppose” is a transitive verb and do not need “to” to specify the object.
- (24) Source: 変革を求める民意は社会の底流に行き続けている。
MT output: The will of the people for change is still there at the undercurrents of society.
Explanation: This should be “in.”
- (25) Source: この薬は、寝る前に飲んでください。
MT output: Drink this pill before going to bed.
Explanation: Unnatural collocation of “drink” and “pill.”

[10] Fluency/Cohesiveness

This type is applied to the cases where the cohesiveness hurts. This include the inappropriate ordering of contents, unclear coreferential mentions, and so forth. If an omission deteriorates the cohesiveness of translation, the corresponding parts in the source side is annotated even if the proper meaning can be conveyed with the omission.

- (26) Source: 京都には世界中から観光客が集まります。多くは、古き良き日本の雰囲気求めてやってきます。
MT output: Kyoto attracts foreign visitors from all over the world. Many come to seek the ancient atmosphere of Japan.
Explanation: It is unclear that this indicates a part of “visitors.”
- (27) Source: こちらの用紙に名前と住所をご記入ください。また、住所には郵便番号も忘れずにご記入ください。
MT output: Please write your name and address on this form. Also, please make sure to include the zip code in your address.
Explanation: The second sentence does not request a new thing but adds a small note; “Also” is not necessary.
- (28) Source: デルタ航空は ... デルタ航空はまた ...
MT output: Delta Air Lines introduces ... Delta Air Lines is also starting ...
Explanation: Verbose repetition of the same subject. Short form or pronouns should be a better device.

[11] Fluency/Register

Register issues include the translations that do not fit the register. For instance, informal expressions and pronouns can be considered as this type of issue when their formal counterparts are required.

- (29) Source: ご不明な点があれば窓口の者にご相談ください。

MT output: Please consult with to the guys at the window if you have any question.

Explanation: The expression is too broken.

- (30) Source: 書類はできるだけ早く提出してください。

MT output: Please submit the documents ASAP.

Explanation: In formal document, depending on the context (cf. evacuation order), abbreviation should not be used.

- (31) Source: 提案手法は有効ではなかった。

MT output: The proposed method isn't effective.

Explanation: In academic article, contractions are not used.

[12] Fluency/Inconsistency

Translations of non-terminological words and phrases are inconsistent. Note that inconsistency of spelling variants is carried forward to Style/Awkward.

- (32) Source: 本装置は、...。当該装置は、...。

MT output: This apparatus comprises, The device improves ...

Explanation: The same referent is translated into different expressions.

- (33) Source: 例えば...。例えば...。

MT output: For example, ... For instance, ...

Explanation: Inconsistent use of typical functional phrases.

[13] Fluency/Others

Fluency of translation is degraded due to other reasons than the above pre-determined reasons.

Style

This top category evaluates the style and notation of the translation. This category does not cover personal preferences and other phenomena that cannot be objectively explained.

[14] Style/Awkward

No grammatical errors, but the translation contains some stylistic issues.

- (34) Source: 図 2 は... を表す。図 3 は... を表す。

MT output: Figure 2 shows ... Fig. 3 demonstrates that ...

Explanation: Spelling variants are not consistently used.

- (35) Source: ... カスタマイズし、... コンテンツをローカライズする。

MT output: ... customise the functions ... localize the content ...

Explanation: Both British and American spellings are used. Note that this does not hurt fluency in terms of grammaticality of sentences (Fluency/Spelling).

Locale

This top category evaluates if the translation fulfills the norms and conventions used in the sphere of the target language.

[15] Locale/Format

Issues of formatting title, address, currency, date, time, measure, number, phone number, etc.

- (36) Source: AAA 国の民主革命から 3 年が経過した。
MT output: It has been 3 years since the democratic revolution happened in AAA.
Explanation: Small numbers must be spelled out.
- (37) Source: ... が 23 日に発覚した。
MT output: ..., sources said on 23.
Explanation: While the date is used in Japanese articles, day of the week is used in English.
- (38) Source: 平成 13 年に公開された千と千尋の神隠しは宮崎駿の代表作だ。
MT output: Spirited Away released in Heisei thirteen is a representative work of Hayao Miyazaki.
Explanation: Era names in Japanese must be converted into Christian era.

[16] Source error

The source contains expressions that cannot be properly translated.

[17] Non-translation

The whole sentence is not relevant to the source sentence, or it is hard to distinguish individual issues.

[18] Other errors

Other types issues.