Guide to selecting MQM issues for the MT Evaluation Metric

version 1.3 (2014 June 11)

Selecting issues can be a complex task. In order to assist annotators, the QTLaunchPad project has prepared a decision tree that helps annotators select appropriate issues. **Use the decision tree not only for learning about MQM issues, but to guide your annotation efforts and resolve any questions or concerns you may have.**

Start at the upper left corner of the decision tree and then answer the questions and follow the arrows to find appropriate issues. The decision tree is organized a bit differently than the hierarchy in translate5 because it eliminates specific issue types before moving to general ones, so you familiarize yourself with how issues are organized in translate5 before beginning annotation.

Add notes in translate5 to explain any decisions that you feel need clarification, to ask questions, or to provide information needed to understand issues, such as notes about what has been omitted in a translation.

In addition to using the decision tree, please understand and follow the guidelines in this document. Email us at *info@qt21.eu* if you have questions that the decision tree and other content in this document do not address.

1. What is an error?

An error represents any issue you may find with the translated text that either does not correspond to the source or is considered incorrect in the target language. The list of language issues upon which you are to base your annotation is described in detail below and provides a range of examples.

The list is divided into two main issue categories, **Accuracy** and **Fluency**, each of which contains relevant, more detailed subcategories. Whenever possible, the correct subcategory should be chosen; however, if in doubt, please do not guess. Instead, select the category level about which you are most certain in order to avoid inconsistencies in the results.
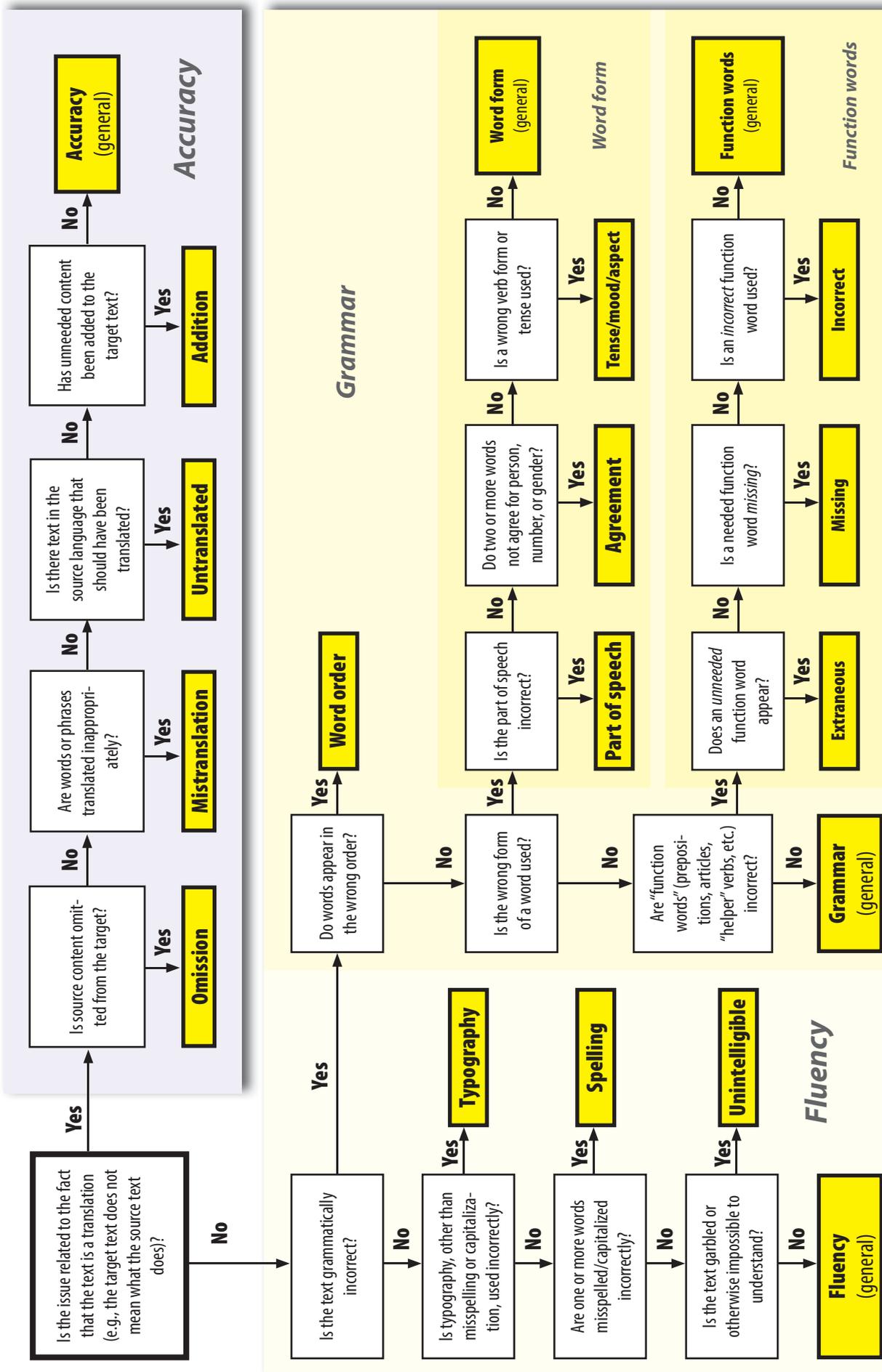
> **Example**: The German term *Zoomfaktor* was incorrectly translated as *zoom shot factor*, and you are unsure whether this represents a **Mistranslation** or an **Addition**. In this case, categorize the error as an **Accuracy** error since it is unclear whether content has been added or a term mistranslated.

2. The Annotation Process

The translations you annotate should be a set of "near miss" (i.e., "almost perfect") translations to annotate. Please follow these rules when selecting errors and tagging the respective text in the translations:

1. Use the examples in this documentation to understand specific classes.

2. If multiple types could be used to describe an issue (e.g., **Agreement**, **Word form**, **Grammar**, and **Fluency**), select the first one that the decision tree guides you to. The tree is organized along the following principles:

   a. It prefers more specific types (e.g., **Part of speech**) to general ones (e.g., **Grammar**). However, if a specific type does not apply, it guides you to use the general type.

   b. General types are used where the problem is of a general nature *or where the specific problem does not have a precise type*. For example *He slept the baby* exhibits what is technically known as a *valency error*, but because there is no specific type for this error available, it is assigned to **Grammar**.

3. **Less is more**. Only tag the relevant text. For example, if a single word is wrong in a phrase, tag only the single word rather than the entire phrase. If two words, separated by other words, constitute an error, mark only those two words separately. (See the section on "minimal markup" below.)

4. If correcting one error would take care of others, tag *only* that error. For example, if fixing an **Agreement** error would fix other related issues that derive from it, tag only the **Agreement** error, not the errors that result from it.

Note: For any question, if the answer is unclear, select "No"

Is the issue related to the fact that the text is a translation (e.g., the target text does not mean what the source text does)?

- **Yes** →

Accuracy

- Is source content omitted from the target? — **Yes** → **Omission** / **No** →
- Are words or phrases translated inappropriately? — **Yes** → **Mistranslation** / **No** →
- Is there text in the source language that should have been translated? — **Yes** → **Untranslated** / **No** →
- Has unneeded content been added to the target text? — **Yes** → **Addition** / **No** → **Accuracy (general)**

- **No** →

Fluency

- Is the text grammatically incorrect? — **Yes** → (to Grammar) / **No** →
- Is typography, other than misspelling or capitalization, used incorrectly? — **Yes** → **Typography** / **No** →
- Are one or more words misspelled/capitalized incorrectly? — **Yes** → **Spelling** / **No** →
- Is the text garbled or otherwise impossible to understand? — **Yes** → **Unintelligible** / **No** → **Fluency (general)**

Grammar

- Do words appear in the wrong order? — **Yes** → **Word order** / **No** →
- Is the wrong form of a word used? — **Yes** → (to Word form) / **No** →
- Are "function words" (prepositions, articles, "helper" verbs, etc.) incorrect? — **Yes** → (to Function words) / **No** → **Grammar (general)**

Word form

- Is the part of speech incorrect? — **Yes** → **Part of speech** / **No** →
- Do two or more words not agree for person, number, or gender? — **Yes** → **Agreement** / **No** →
- Is a wrong verb form or tense used? — **Yes** → **Tense/mood/aspect** / **No** → **Word form (general)**

Function words

- Does an *unneeded* function word appear? — **Yes** → **Extraneous** / **No** →
- Is a needed function word *missing*? — **Yes** → **Missing** / **No** →
- Is an *incorrect* function word used? — **Yes** → **Incorrect** / **No** → **Function words (general)**

Examples

Source: Importfilter werden geladen
Translation: Import filter are being loaded
Correct: Import filters are being loaded

In this example, the only error is the translation of *filter* in the singular rather than the plural (as made clear by the verb form in the source text). This case should be classified as **Mistranslation**, even though it shows problems with agreement: if the subject had been translated properly the agreement problem would be resolved. In this case only *filter* should be tagged as a **Mistranslation**.

Source: im Dialog Exportieren
Translation: in the dialog export
Correct: in the Export dialog

In this example, only **Mistranslation** should be marked. While **Word order** and **Spelling** (capitalization) would be considered errors in other contexts, this would not be the case here, as these two words constitute one term that has been incorrectly translated.

5. If one word contains two errors (e.g., it has a **Spelling** issue and is also an **Extraneous function word**), enter both errors separately and mark the respective word in both cases.

6. If in doubt, choose a more general category. The categories **Accuracy** and **Fluency** can be used if the nature of an error is unclear. In such cases, providing notes to explain the problem will assist the QTLaunchPad team in its research.

3. Tricky cases
The following examples are ones that have been encountered in practice and that we wish to clarify.

- **Function words**: In some cases issues related to function words break the accuracy/fluency division seen in the decision tree because they are listed under *Fluency* even though they may impact meaning. Despite this issue, please categorize them as the appropriate class under *Function words.*

  Example: *The ejector may be found **with** the external case* (should be **on** in this case). Even though this error changes the meaning, it should be classified as **Function words: incorrect** in the **Fluency** branch.

- **Word order**: Word order problems often affect long spans of text. When encountering word orders, mark the smallest possible portion that could be moved to correct the problem.

  Example: *He has the man with the telescope <u>seen</u>.* Here only *seen* should be marked as moving this one word would fix the problem.

- **Hyphenation**: Hyphenation issues sometimes occur in untranslated content and should be classified as such. Otherwise they should be classified as **Spelling**.

  Example: *Load the XML-files* (**Spelling**)
  Nützen Sie die macro-lens (**Untranslated**, if the source has *macro-lens* as well)

- **Number** (plural vs. singular) is a **Mistranslation**.

- **Terminology**: Inappropriate use of terms is classified as **Mistranslation**.

  Example: An English translation uses the term *thumb drive* to translate the German *USB Speicherkarte*. This translation is intelligible, but if the translation mandated in specifications or a relevant termbase is *USB memory stick*, the use of *thumb drive* constitutes a **Mistranslation**, even if *thumb drive* would be acceptable in everyday usage.

- **Unintelligible**: Use **Unintelligible** if content cannot be understood and the reason cannot be analyzed according to the decision tree. *This category is used as a last resort for text where the nature of the problem is not clear at all.*

    Example: In the sentence "You can also you can use this tab to precision, with the colours are described as well as the PostScript Level," there are enough errors that the meaning is unclear and the precise nature of the errors that lead to its unintelligibility cannot be easily determined.

- **Agreement**: This category generally refers to agreement between subject and predicate or gender and case.

    Examples: The boy was playing with *her* own train
    I *is* at work

- **Untranslated**: Many words may look as if they have been translated and simply forgotten to apply proper capitalization or hyphenations rules. In most, cases, this would represent an untranslated term and not a **Spelling**. If the target word or phrase is identical to the source word or phrase, it should be treated as **Untranslated**, even if a **Spelling** error could also account for the problem.

4. Minimal markup

It is vital in creating error markup that errors be marked up with the shortest possible spans. Markup must identify *only* that area needed to specify the problem. In some cases this requirement means that two separate spans must be identified.

The following examples help clarify the general principles:

Incorrect markup	Problem	Correct minimal markup
Double click on the number faded in the status bar. [**Mistranslation**]	Only the single word *faded* is problematic, but the markup indicates that number faded in is incorrect.	Double click on the number faded in the status bar.
The standard font size for dialogs is 12pt, which corresponds to a standard of 100%. [**Mistranslation**]	Only the term *Maßstab* has been translated incorrectly. The larger span indicates that text that is perfectly fine has a problem.	The standard font size for dialogs is 12pt, which corresponds to a standard of 100%.
The in 1938 nascent leader with flair divined %temp_name eating lonely. [**Unintelligible**]	The entire sentence is **Unintelligible** and should be marked as such.	The in 1938 nascent leader with flair divined %temp_ name eating lonely.

As noted above, **Word order** can be problematic because it is often unclear what portion(s) of the text should be marked. In cases of word order, mark the *shortest* portion of text (in number of words) that could be moved to fix the problem. If two portions of the text could resolve the problem and are equal in length, mark the one that occurs first in the text. The following examples provide guidance:

Incorrect markup	Problem	Correct minimal markup
The telescope big observed the operation	Moving the word *telescope* would solve the problem and only this word should be marked (since it occurs first in the text).	The telescope big observed the operation
The eruption by many instruments was recorded.	Although this entire portion shows word order problems, moving *was recorded* would resolve the problem (and is the shortest span that would resolve the problem).	The eruption by many instruments was recorded.
The given policy in the manual user states that this action voids the warranty.	This example actually has two separate issues that should be marked separately.	The given policy in the manual user states that this action voids the warranty.

Agreement poses special challenges because portions that disagree may be widely separated. To select appropriate minimal spans, consider the following guidelines:

- If two items disagree and it is readily apparent which should be fixed, mark *only* the portion that needs to be fixed. E.g., in "The man and its companion were business partners" it is readily apparent that *its* should be *his* and the wrong grammatical gender has been used, so only *its* should be marked.

- If two items disagree and it is not clear which portion is incorrect, mark the both items and mark them for **Agreement**, as shown in the example in the table below.

The following examples demonstrate how to mark **Agreement**:

Incorrect markup	Problem	Correct minimal markup
The man and its companion were business partners. [**Agreement**]	In this example, it is clear that *its* is the problematic portion, and that *man* is correct, so only *its* should be marked.	The man and its companion were business partners.
The man whom they saw on Friday night at the store were very big. [**Agreement**]	In this example it is not clear whether *man* or *were* is the error since there is nothing to indicate whether singular or plural is intended. Here the highlighted portion identifies only a single word, insufficient to identify the agreement problem. The correct version highlights both words as separate issues. In such cases use the Notes field to explain the decision.	The man whom they saw on Friday night at the store were very big. [**Agreement**]

In the event of questions about the scope of markup that should be used, utilize the Notes field to make a query or explain your choice.

A. Issue categories

The error corpus uses the following issue categories:

- **Accuracy**. **Accuracy** addresses the extent to which the target text accurately renders the meaning of the source text. For example, if a translated text tells the user to push a button when the source tell the user not to push it, there is an accuracy issue.

  - **Mistranslation**. The target content does not accurately represent the source content.

    Example: A source text states that a medicine should not be administered in doses greater than 200 mg, but the translation states that it should *not* be administered in doses less than 200 mg.

    Note(s): **Mistranslation** can be used for both words and phrases.

  - **Omission**. Content is missing from the translation that is present in the source.

    Example: A source text refers to a "mouse pointer" but the translation does not mention it.

    Note(s): **Omission** should be reserved for those cases where content present in the source and essential to its meaning is not found in the target text.

  - **Addition**. The target text includes text not present in the source.

    Example: A translation includes portions of another translation that were inadvertently pasted into the document.

  - **Untranslated**. Content that should have been translated has been left untranslated.

    Example: A sentence in a Japanese document translated into English is left in Japanese.

    Note(s): As noted above, if a term is passed through untranslated, it should be classified as **Untranslated** rather than as **Mistranslation**.

- **Fluency**. Fluency relates to the monolingual qualities of the source or target text, relative to agreed-upon specifications, but independent of relationship between source and target. In other words, fluency issues can be assessed without regard to whether the text is a translation or not. For example, a spelling error or a problem with register remain issues regardless of whether the text is translated or not.

  - **Spelling**. Issues related to spelling of words (including capitalization)

    Examples: The German word *Zustellung* is spelled *Zustetlugn*.
    The name *John Smith* is written as "john smith".

  - **Typography**. Issues related to the mechanical presentation of text. This category should be used for any typographical errors other than spelling.

    Examples: Extra, unneeded carriage returns are present in a text.
    A semicolon is used in place of a comma.

  - **Grammar**. Issues related to the grammar or syntax of the text, other than spelling and orthography.

    Example: An English text reads "The man was in seeing the his wife."

    Note(s): Use **Grammar** only if no subtype accurately describes the issue.

    - **Word form**. The wrong form of a word is used. Subtypes should be used when possible.

      Example: An English text has *comed* instead of *came*.

      - **Part of speech. A word is the wrong part of speech**

        Example: A text reads "Read these instructions *careful*" instead of "Read these instructions *carefully*."

- **Agreement. Two or more words do not agree with respect to case, number, person, or other grammatical features**

    Example: A text reads "*They was* expecting a report."

- **Tense/aspect/mood. A verbal form inappropriate for the context is used**

    Example: An English text reads "Yesterday he *sees* his friend" instead of "Yesterday he *saw* his friend"; an English text reads "The button must be *pressing*" instead of "The button must be *pressed*".

- **Word order**. The word order is incorrect

    Example: A German text reads "Er hat gesehen den Mann" instead of "Er hat den Mann gesehen."

- **Function words**. Linguistic function words such as prepositions, particles, and pronouns are used incorrectly

    Example: An English text reads "He beat him around" instead of "he beat him up."

    Note(s): Function words is used for cases where individual words with a grammatical function are used incorrectly. The most common problems will have to do with prepositions, and particles. For languages where verbal prefixes play a significant role in meaning (as in German), they should be included here, even if they are not independent words.

    There are three subtypes of *Function words*. These are used to indicate whether an unneeded function word is present (**Extraneous**), a needed function word is missing (**Missing**), or a incorrect function word is used (**Incorrect**). Annotators should use the note field to specify details for missing function words.

- **Unintelligible**. The exact nature of the error cannot be determined. Indicates a major break down in fluency.

    Example: The following text appears in an English translation of a German automotive manual: "The brake from whe this કુતરો સિ S149235 part numbr,,."

    Note(s): Use this category sparingly for cases where further analysis is too uncertain to be useful. If an issue is categorized as **Unintelligible** no further categorization is required. Unintelligible can refer to texts where a significant number of issues combine to create a text for which no further determination of error type can be made or where the relationship of target to source is entirely unclear.