Guidelines for Error Annotation

1. Introduction

Your task is to annotate errors in the translations produced by four different machine translation systems. Annotation will be carried out in the *Accolé* web platform. The dataset to be annotated consists of 200 segments; each segment is made up of one source sentence and four different target sentences (one target per system). In addition, each segment comes with a reference translation produced by a human translator. Thus, one segment comprises a total of six sentences: one source sentence, one human reference translation, and the four machine-generated target sentences to be annotated.

In *Accolé*, you will annotate one segment after the other. For each segment, all six sentences are displayed on one page. The annotation is straightforward: read the source and target sentences carefully, and if you find an error or problem in a given target sentence, use the mouse to mark the problematic word or sequence of words. In addition, mark the corresponding words in the source sentence. Once the problematic target words and the corresponding source words are marked, select an error type from the dropdown menu. After having annotated all identified errors in this way, proceed to the next segment.

The main aim of this study is not to assess the quality of machine translation output as compared to human translations, but to conduct a fair comparison between four different machine translation systems. So please try to avoid human-versus-machine attitudes; instead, imagine you are giving feedback to four different students to help them improve their translation skills. In such a task, it is essential to be as consistent as possible in your error judgments.

Detecting and categorizing errors can be a complex task. In order to assist you, this document contains a description of the error typology (Section 3), a decision tree that helps select appropriate error types (Section 3.3), and a description of the annotation process (Section 4). In the end, the goal of the task is to produce a consistent error annotation, which means that you should try not to select errors according to your intuition, but according to the specifications provided in the present guidelines.

2. What is an Error?

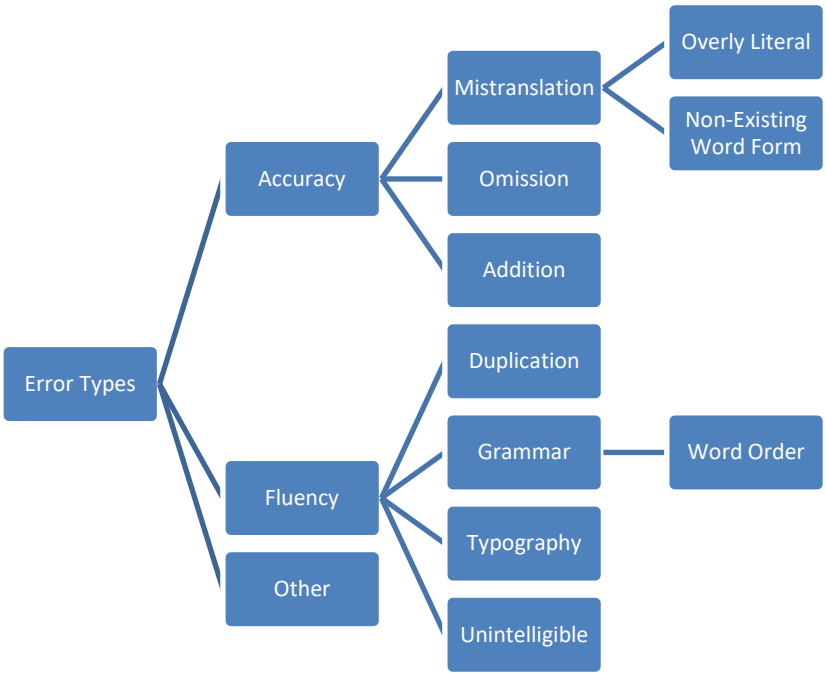
An error represents any problem you may find in the translated texts that either does not correspond to the source or that is considered incorrect in the target language. The typology of errors upon which you are to base your annotation is described in Section 3 and provides a range of examples.

The error typology is divided into two main error categories: **Accuracy**, which refers to the transfer of meaning from the source to the target language, and **Fluency**, which refers to the extent translations conform to target language norms. Both categories are further subdivided into more detailed subcategories. Whenever appropriate, the subcategory should be chosen.

You may encounter borderline cases for which it will be difficult to decide whether there is an error in the translation. For example, think of issues such as style or idiomaticity, where the lines between error and non-error are blurred. For cases of unidiomatic wording, mark them as fluency errors (or, if applicable, a more detailed subtype) if they have a strong impact on the readability or intelligibility of the translation. If the impact is subtle or negligible (e.g. if you have to be a language expert to spot the issue), ignore them. The same applies to stylistic issues, for example mismatches of register or tone between the source and the target, such as the use of formal register in the source versus informal or colloquial wording in the target. If the mismatch strongly alters the meaning conveyed by the translation, an accuracy error – or, if applicable, a more detailed subtype – should be marked, if the difference is only nuanced, ignore it. If in doubt, always mark an error rather than ignoring it. Most importantly, try to be consistent your judgments: please try to be equally strict not only for all sentences within a given segment, but also across all other segments of the dataset.

3. Error Typology

3.1. Overview of Error Types



3.2. Description of Error Types

3.2.1. Accuracy

Accuracy addresses the extent to which the target accurately **transfers the meaning of the source** text. Such errors can be detected **only by comparing the target with the source**. For example, if a translated text tells the user to push a button when the source tells the user not to push it, there is an accuracy issue. Accuracy errors are further subdivided into the following more detailed subcategories:

- **Mistranslation**: Words or phrases in the target do not accurately represent the meaning of the source content, i.e. words or phrases were translated incorrectly.

  Example: A source text states that the name of a given organization '*dates back* to the year 1920', but the translation states that the organization's name '*was used* in the year 1920' (without mentioning that in 1920 it was used for the first time).

  Example: The English translation states that a political party had '*hardly any* concerns', whereas the source states that there were '*persistent* concerns'.

  Example: Mismatch of number between source and target: in the source, 'the *documents are* uploaded' (plural), whereas in the target 'the *document is* uploaded' (singular).

  - **Overly literal**: The translation is too literal, thus obscuring the meaning of a word or phrase conveyed by the source. This is often the case with idiomatic expressions or phrasal verbs translated word for word into the target language.

    Example: The German source states that a person was 'oben ohne', which means 'top-less' in English. The English target states that the person was 'above with-out', which is a meaningless word-for-word translation.

    Example: The German phrase 'Es geht.', used as a reply to the question 'Wie geht's?' (meaning 'How are you?'), is translated literally as 'It is going' rather than the equivalent 'Not too bad' or 'So-so'.

□ **Non-existing word form**: The translation contains a word that does not exist in the target language, thus obscuring or distorting the meaning conveyed by the source. The non-existing word is an invention (hallucination) of the machine translation system.

Example: The English target mentions '*cristious* concerns'. Due to the non-existing adjective 'cristious', the accuracy of the translation cannot be determined.

Example: The German word 'Donnerstag' (meaning 'Thursday'), is translated as 'Donnerday'. Or the German 'Knappheit' (meaning 'scarcity'), is translated as 'nappyness'. Both instances are non-existing words made up by the system.

Note: In some cases, the **line between non-existing words forms and spelling errors is blurred**. You are asked to mark spelling errors, i.e. unwarranted deviations of orthographic rules that may or may not hinder comprehension, as non-existing word form errors. For example, misspellings of the word 'narcissism' as 'narcissm' or of the word 'secondary' as 'sekundera' are to be marked as non-existing word form errors.

Note: When you identify a non-existing word form, it is important to keep in mind the distinction between accuracy errors and fluency errors. In some cases, non-existing word forms may be instances of fluency errors. Example: the German past tense form of the verb 'bekommen' (meaning 'to receive, get') is rendered into 'bekommte' instead of the correct 'bekam'. Since the lexical meaning of the source is correctly transferred, this non-existing inflectional form should be marked as *Grammar* error.

Note: Compounds that cannot be found in a dictionary but that fully correspond to the word formation rules of the target language are *not* to be marked as errors. For example, the English 'twelve-day' (as in 'twelve-day period') or the German 'Tassenmiete' (meaning 'cup rental') are infrequent yet likely and correct creative recombinations of existing lexical material. Therefore, no error is to be marked.

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

Example: In the source, a statement is negated, for instance 'I *don't* want to go home', whereas in the target the negation is missing, e.g. 'I want to go home'.

Example: A source text refers to an event and its date, but the translation mentions only the event without specifying the date it took place.

Note: Use *Omission* only for those cases where content present in the source and essential to its *meaning* is not found in the target. In case of missing function words (e.g. articles, prepositions, etc.), mark the omission as *Grammar* error.

Note: Mark the omission in the target at the position where the missing word should appear by selecting the gap between the two words before and after the omission. For example, if the adjective 'big' is missing from 'I see a house', select the gap between 'a' and 'house' to mark the omission. The selection of gaps does not work in Firefox, so check whether your browser supports this function. Whenever possible, also try to mark the source words or word sequences that are missing from the target. While this is straightforward for content words, it may be impossible for fluency-related omissions, e.g. missing function words.

Note: There are no clear-cut boundaries between omissions and implicitation. Implicitation is a legitimate translation strategy and the result of excluding content that can be inferred from the context. Omissions, on the other hand, are the result of excluding content that is not, in any way, inferable from the context. When in doubt, you are asked to mark an *Omission* error. Example: The German

source word '*Land*straße' (meaning 'country road') is translated as 'road'. Irrespective of whether it can be inferred from the context that the road is located in the countryside, you should mark an *Omission* error.

Note: If longer source passages, e.g. entire clauses, are missing from the target, try to mark as many omissions as there are missing phrases (see also Section 4.3.2).

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

  Example: The source says that someone wants to go home, whereas in the target a negation has been added (e.g. 'I *don't* want to go home'), thus altering the meaning.

  Note: Instances of one or more superfluous function words (e.g. articles, prepositions) should be marked as *Grammar* errors. Thus, *Addition* is reserved for additions of single content words, or phrases that contain at least one content word.

  Note: There are no clear-cut boundaries between additions and explicitation. Explicitation is a legitimate translation strategy and the result of explicitly mentioning content in the target that is only implicitly encoded in the source. Additions, on the other hand, are the result of encoding content that is neither explicitly nor implicitly mentioned in the source. When in doubt, you are asked to mark an *Addition* error. Example: The German source 'führende Forscher entwickelten' was translated as 'leading researchers *in the field* developed'. Although the added 'in the field' is inferable from the context, please mark an *Addition* error.

  Note: The distinction between *Duplication* and *Addition* errors may be difficult at times. Please use *Addition* if the target contains text that is not contained in the meaning of the source. Duplications, on the other hand, refer to target words or phrases that do have a semantically equivalent counterpart in the source and which appear repeatedly in the target while the source contains only one instance. When in doubt whether, always mark *Duplication*.

3.2.2. Fluency

Fluency relates to the mechanics of the target language. Fluency errors violate target language norms and can be identified **without reading the source text**.

- **Duplication**: One or more target words have been duplicated, i.e. they are repeated unwarrantedly.

  Example: A translation reads 'for example, there was a problem, *for example'*, whereas the source contains only one equivalent of 'for example'.

  Note: According to the markup principle *Less is More* (see 4.3.1), in the target only the second instance of the duplicated word rather than both instances should be annotated. In the above example, only the second instance of 'for example' should be marked. Do not forget to mark the duplicated words in the source as well.

  Note: The distinction between *Duplication* and *Addition* errors may be difficult at times. Please use *Addition* if the target contains text that is not contained in the meaning of the source. Duplications, on the other hand, refer to target words or phrases that do have a semantically equivalent counterpart in the source and which appear repeatedly in the target while the source contains only one instance. When in doubt whether, always mark *Duplication*.

- **Typography**: The target text violates the typographic norms of the target language (punctuation, spacing, incorrect letter case, etc).

  *Example***:** A semicolon is used in place of a comma.

*Example***:** Incorrect letter case, e.g. the proper noun 'John' is written as '*j*ohn'.

Example: One of a pair of quotation marks or brackets is missing from the target.

Note: English compounds can be written as open compounds (spelled as two words, e.g. 'ice cream'), closed compounds (spelled as one word, e.g. 'goldfish'), or hyphenated compounds (e.g. 'long-term'). In German, mostly closed compounds are used (e.g. 'Goldfisch'). Spelling errors in compounds resulting from missing or superfluous whitespaces or hyphens should be marked as *Typography* errors. Examples: 'gold fish' instead of 'goldfish', 'icecream' instead of 'ice cream', or 'long term' instead of 'long-term' in English; or 'Gold Fisch' instead of 'Goldfisch' in German.

- **Grammar:** This is a rather broad category that encompasses problems related to the grammar (syntax or morphology) of the target text, other than spelling and typography. Such errors include lack of agreement between subject and verb, word order, incorrect declension of nouns, etc. Function words may also be the source of errors, e.g. missing, superfluous or incorrect function words.

  Example: An English text reads 'The man was seeing *the his* wife.'

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

  - **Word order**: Target words appear in the wrong order.

    Example: An English text reads 'He the car *saw'* instead of 'He *saw* the car'.

    Note: Word order errors often affect long spans of text. When encountering word order errors, mark the smallest possible portion of text that could be moved to correct the problem. For example, in 'He has the man with the hat *seen'*, only the word 'seen' should be marked, because moving this one word would fix the problem (see 4.3.3).

    Note: If two portions of the text could resolve the problem and both are equal in length, mark the one that occurs first in the text. For example, in 'the dog big', moving either 'dog' to the third or 'big' to the second position would fix the problem. Since 'dog' appears first in the target text, only this word should be annotated as an error.

- **Unintelligible**: The target content cannot be understood due to a major break down in fluency; therefore, the exact nature of the fluency error cannot be determined.

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

  Note: This category is used **only as a last resort** for text where the nature of the fluency problem is not clear at all. Use this category **sparingly** for cases where further analysis is too uncertain to be useful.

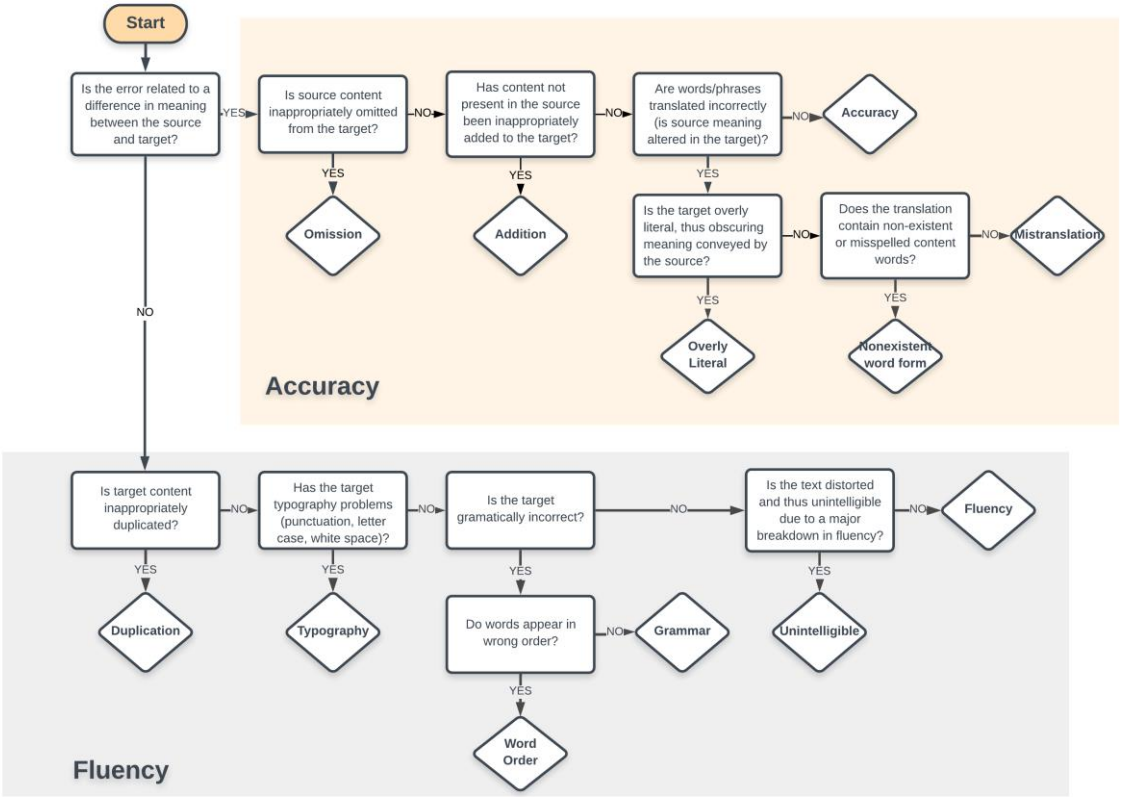
  Note: *Unintelligible* can refer to texts where a large number of errors combine to create a text for which no further error analysis can be made or where the relationship of target to source is entirely unclear. For example, in 'The in 1938 nascent leader with flair divined eating lonely' there are so many errors that the meaning of entire sentence is unclear. Therefore, the *entire* sentence should be marked as *Unintelligible*.

3.2.3. Other

This error type is used for any errors not adequately covered by any of the error types described above. This category should be used ***only*** if it is impossible to assign an error to an existing category with sufficient granularity. In other words: *Other* is to be used ***only*** when it is entirely unclear whether an error relates to *Accuracy* or *Fluency*.

3.3. Decision Tree

Use the decision tree not only to familiarize yourself with the error types, but to guide your annotation efforts and to resolve any questions or concerns you may have.
When using the decision tree, start at the upper left corner of the decision tree and then answer the questions and follow the arrows to find appropriate error types. A high-resolution PDF of the decision tree is attached to these guidelines.

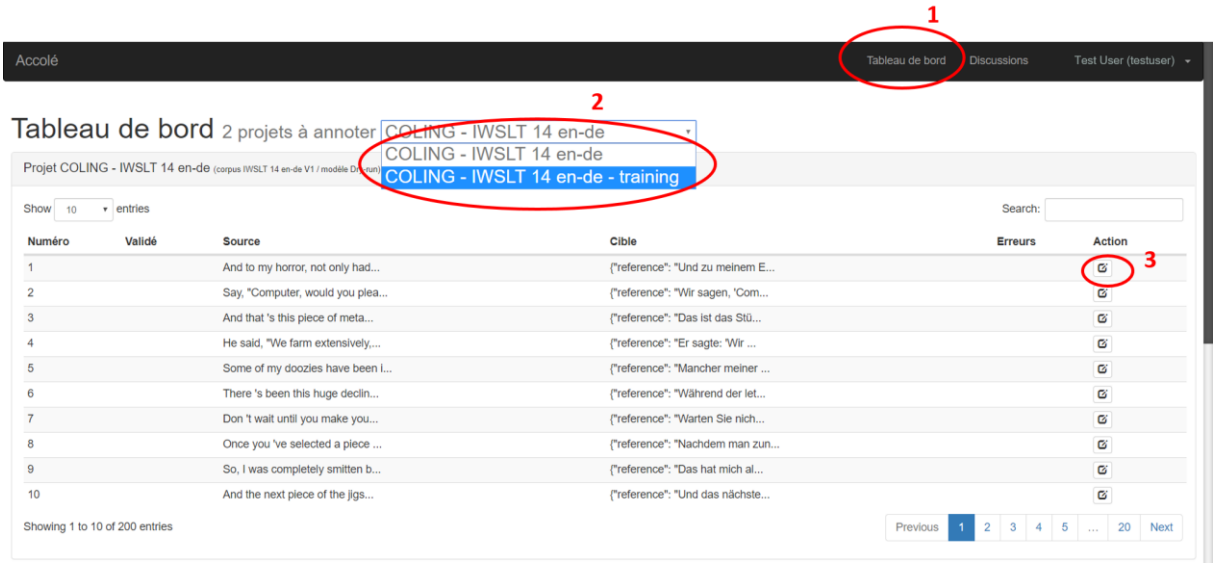


4. The Annotation Process

4.1. The *Accolé* Error Annotation Interface

4.1.1. Login and Start Annotation

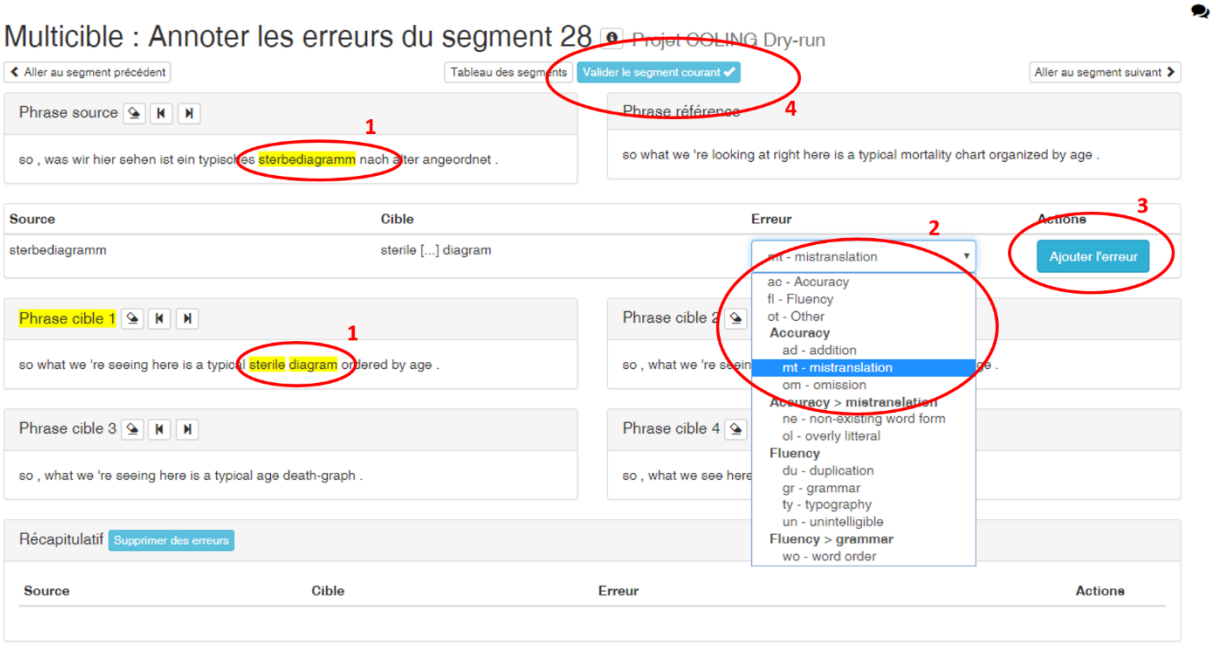
Please go to the login page (http://lig-accole.imag.fr/app.php/login) and enter your personal login credentials. To start the annotation, proceed as follows (see also Screenshot 1 below):



1. Click on the tab “Tableau du bord” to show the overview of all projects assigned to you.
2. Choose the project from the dropdown menu.
3. In the list of all segments within a given project, click the button  in the Action column of the first segment. This will bring you to the annotation window (see 4.1.2).

4.1.2. Conducting the Annotation

Error annotation consists of a) marking (= highlighting) it in the source and the target, and b) assigning an error type to it. To annotate all errors for a given segment in *Accolé*, please proceed as follows:



- **Mark the target word(s)** that represent an error in the target sentence by highlighting the word(s) with the mouse pointer (Step 2 in the above screenshot).

  Multiple words: To mark one word, just click on the word with the left mouse button. The mark a sequence of multiple adjacent words, left-click on the first word (does not have to be the beginning of the word), hold down the mouse button and move the cursor to the last word of the sequence (does not have to be the end of the word) before releasing the mouse button. To mark multiple non-adjacent words (e.g. to mark both parts of German split verbs, such as 'Peter wachte um 8 Uhr auf'), just click on each word one by one.

  Omission errors: To mark omission errors, select the gap between the words where the omitted element should appear. This feature is not available in Firefox.

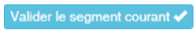
  Undo markup: To undo or delete the markup, click on the erase button  .

  Mark start/end: To mark the very first or very last position of the sentence (e.g. to indicate that a word is missing in the beginning of the sentence or that a full stop is missing at the end of the sentence), click the button  or  , respectively.

- In the same way, **mark those source word**(s) that correspond to the incorrect target words (Step 1). Sometimes it is not possible or meaningful to mark source words, for example in the case of addition errors (target content that is not present in the source).

- After marking the target and, where applicable, the source words, **select the error type** from the dropdown menu (Step 2). Subsequently, click on the button *Ajouter l'Erreur* (Step 3) to save the error. All saved errors are summarized at the bottom of the page in the *Récapitulatif* table.

  Keyboard shortcut: Instead of selecting an error type from the dropdown with the mouse, you can type the two-letter code of an error after having marked the span of the error. Note: If you select the error in this way, you also need to click *Ajouter l'Erreur* in order to save the error.

- After having marked and error-coded all errors in all target sentences, click on the blue button ***Valider le segment courant*** 

(Step 4) to confirm your annotation for the given segment and to move on to the next segment. To go back to the previous segment, click on the button *Aller au segment précédent* at the top left.

- In case you want to comment on a given segment (e.g. to point out that you did not understand the source sentence, or that you are absolutely uncertain about your annotation for the given segment), click on the speech bubble button  (top right). Please use this function sparringly; instead, try to make careful decisions in your annotations by relying on the present guidelines.

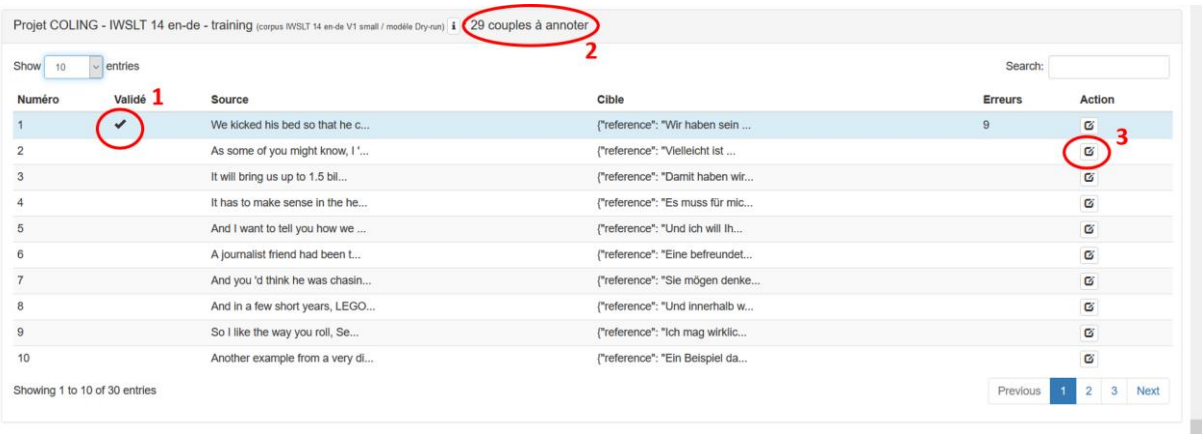
- To delete an already saved error, first click on the blue button *Supprimer des erreurs* in the *Récapitulatif* table at the bottom of the page, then click the red button 

that appears next to the saved error in the rightmost column of the *Récapitulatif* table. There will be no further warning, so be careful when deleting errors.

Please note: The human **reference translation is provided *only for guidance***, for example if you do not understand the source. Your error annotation should not be based on the wording of the human reference translation, but on the relationship between the source and each target sentence.

4.1.3. Resume Annotation

To resume the annotation of a project that you have not yet completed, please follow these steps (Screenshot 3 below):



1. After logging into *Accolé*, click on the *Tableau du bord* tab.
2. In the *Validé* column, all segments that have been fully annotated and validated by clicking Valider le segment courant ✔ are checked (see Nr 1 in Screenshot 3). In the header of the project, the number of remaining segments to be annotated is displayed (Nr 2).
3. Click on the button for the segment you wish to annotate. This will bring you to the annotation window (see 4.1.2).

4.2. Selecting Error Types

The error typology is divided into two main error categories (Accuracy and Fluency), each of which is subdivided into more fine-grained subcategories. Whenever appropriate, the subcategory should be chosen to annotate an error. However, if you have doubts, 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.

Please follow these rules when selecting errors and tagging the respective text in the translations:

1. Use the examples in the present guidelines to understand specific classes.
2. If multiple types could be used to describe an issue (e.g., Word order, Grammar, and Fluency), select the first one that the decision tree (see 3.3) guides you to. The tree is organized along the following principles:
    a. It prefers more specific types (e.g. *Word order*) 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 in 'environment friendly', the word 'environment' is the wrong part of speech (should be 'environmentally'), but since this error type is not included in the error typology, the error should be 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. For more examples, see Section 4.3.1.
4. If correcting one error would take care of others, tag only that error.

    Example:
    Source: Importfilter werden geladen
    Target: Import *filter are* being loaded
    Correct: Import filters are being loaded

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

5. If one word contains two errors, annotate both errors separately by marking and error-coding it twice. For example, in 'He *john* saw.', the word 'john' is a word order error (the word should be moved to the last position), and a typography error (the word should be capitalized).
6. If in doubt, choose a more general category. The categories *Accuracy* and *Fluency* can be used whenever the more specific nature of an error cannot be determined.

4.3. Minimal Markup

4.3.1. Less is More

It is crucial to mark errors with the shortest possible spans. Markup must identify *only* the span of words needed to specify the problem. In some cases this requirement means that two or more separate spans must be identified

Incorrect mark-up	**Problem**	**Correct minimal mark-up**
Double click on the number indicates in the status bar. [**Mistranslation**]	Only the single word *indicates* is problematic (should be *indicated*), but the markup wrongly identifies the sequence *number indicated* as incorrect.	Double click on the number indicates **[Mistranslation]** in the status bar.
The font size is 12pt, which corresponds to a standard of 100%. [**Mistranslation**]	The German *Maßstab* 'scale' has been incorrectly translated as *standard*. The markup wrongly indicates that there is an error in 'of 100%' as well. Only the word containing the error should be annotated.	The font size is 12pt, which corresponds to a standard **[Mistranslation]** of 100%.
He john saw. **[Word order]**	The word *john* contains two errors: a word-order error (should be moved to last position), and a typography error (should be capitalized). In this markup, the correct word *saw* is unwarrantedly tagged.	He john **[Word order] [typography]** saw.
Source markup: Er macht **[Mistranslation]** das Fenster zu. Target markup: He opens **[Mistranslation]** the window.	The German split verb *macht […] zu* (meaning 'he closes') was incorrectly translated as 'he opens'. The markup indicated that only *macht* has been translated incorrectly.	Source markup: Er macht das Fenster zu **[Mistranslation]**. Target markup: He opens **[Mistranslation]** the window.

4.3.2. Marking Omissions of Long Sequences

If longer source passages (e.g. entire clauses) are missing from the target, try to decompose the missing content into smaller meaningful units (e.g. phrases) and mark each omitted unit individually rather than annotating one long omission error only. Consider the following example of the German sentence 'Wir leben an einem wunderbaren Ort mit kostenloser Gesundheitsversorgung für alle', which was

translated as ‘We live in a wonderful place’, omitting the entire prepositional phrase ‘mit kostenloser Gesundheitsversorgung für alle‘ (meaning ‘with free health care for everyone’):

Incorrect mark-p	**Problem**	**Correct minimal markup**
Source markup: Wir leben an einem wunderbaren Ort mit kostenloser Gesundheitsversorgung für alle. **[Omission]** Target markup: We live in a wonderful place. **[Omission]**	The entire sequence *mit kostenlsoer Gesundheitsversorgung für alle* is missing from the target, but only one omission has been marked. Instead, try to mark each omitted meaningful sub-sentence unit (e.g. phrase) separately.	Source: Wir leben an einem wunderbaren Ort mit kostenloser Gesundheitsversorgung **[Omission]** für alle **[Omission]**. Target: We live in a wonderful place **[Omission] [Omission]**.

4.3.3. Markup of Word Order Errors

As noted above, Word order can be problematic because it is often unclear what portion of text should be marked. In cases of word order problems, 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:

Incorrect mark-up	**Problem**	**Correct minimal mark-up**
The dog big **[Word Order]** barked.	Moving the word *dog* would solve the problem and only this word should be marked (since it occurs first in the text).	The dog **[Word Order]** big barked.
The eruption by many instruments was recorded. **[Word Order]**	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. **[Word Order]**
The given policy in the manual user **[Word Order]** states that this action voids the warranty.	This example actually has two separate problems: ‘given policy’ instead of ‘policy given’, and ‘manual user’ instead of ‘user manual’. Both errors should be marked separately.	The given **[Word Order]** policy in the manual **[Word Order]** user states that this action voids the warranty.