ACCOLÉ: A Collaborative Platform of Error Annotation for Aligned Corpora
Emmanuelle Esperança-Rodier, Francis Brunet-Manquat, Sophia Eady

To cite this version:
Emmanuelle Esperança-Rodier, Francis Brunet-Manquat, Sophia Eady. ACCOLÉ: A Collaborative Platform of Error Annotation for Aligned Corpora. Translating and the computer 41, Nov 2019, Londres, United Kingdom. hal-02363208

HAL Id: hal-02363208
https://hal.archives-ouvertes.fr/hal-02363208
Submitted on 14 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
ACCOLÉ: A Collaborative Platform of Error Annotation for Aligned Corpora

Emmanuelle Esperança-Rodier
Univ. Grenoble Alpes, CNRS, Grenoble INP*, LIG, 38000 Grenoble, France
Emmanuelle.Esperanca-Rodier@univ-grenoble-alpes.fr

Francis Brunet-Manquat
Univ. Grenoble Alpes, CNRS, Grenoble INP*, LIG, 38000 Grenoble, France
Francis.Brunet-Manquat@univ-grenoble-alpes.fr

Sophia Eady
Univ. Grenoble Alpes, CNRS, Grenoble INP*, LIG, 38000 Grenoble, France
Sophia.Eady@univ-grenoble-alpes.fr

Abstract

This article presents a platform, named ACCOLÉ, for the collaborative annotation of translation errors. ACCOLÉ offers a range of services that allow simplified management of corpora and error typologies, annotation of effective errors, collaborative discussion during annotation, and finally different kinds of search in corpora. ACCOLÉ allows the annotation of translation errors according to built-in error typologies, Vilar's typology or DQF-MQM or uploaded ones, on several annotated corpora of different texts, translated by different Statistical or Neural MT systems. It also help to process the annotated corpora created in order to look for typical error models and patterns, related to a specific MT system. ACCOLÉ currently provides 15 corpora, 19 projects of 134,273 source words and 114,511 target words, and 23,525 annotations. Eventually, we will implement the semi-automatic propagation of found patterns on other corpora to compare the behaviour of the MT systems on different domains, thus providing to the community a wide range of error-annotated bilingual parallel corpora.

1 State of the Art

Translation error analysis is a task that has been developed in recent years in order to improve the machine translation (MT) systems and to qualitatively evaluate these same systems. In addition to the field of MT, translation error analysis is used in particular to improve translation teaching methodologies (Kübler et al., 2016). The integration of MT in translation masters brings a new questioning about the evaluation of MT systems. How do translators feel about different MT systems (Esperança-Rodier et al., 2017)? What tools to use for non-specialists? The answer to these questions requires a linguistic analysis of MT errors in order to be able, firstly, to reveal the typical errors of each type of MT systems, secondly, to correlate these typical errors with the post-editing capability of the users, and thirdly to create corpora. Performed manually, the translation error analysis is a very time-consuming task. Research is therefore concerned with the automatic analysis of translation errors (Popović et al., 2006). Manual or automatic, this task requires error analysis criteria, often related to the particular characteristics of the tool used.

The work of Vilar et al. (2006), the McLLANGE project (Castagnoli et al., 2011), Felice and Specia (2012), Wisniewski et al. (2013a) and the work of Lommel and Melby (2018) focuses on the definition of typologies of linguistic errors. An error typology consists of a list of errors, categorized according to the corrective actions to be performed to improve the translation. For example, the error type "missing word" indicates that a word of the source segment has not been translated in the translation hypothesis and therefore does not appear there. In order to improve the translation hypothesis, corrective actions must be taken to translate this "missing word". ACCOLÉ proposes these different typologies of error

* Institute of Engineering Univ. Grenoble Alpes
annotation, but also makes it possible to propose new typologies, better adapted to the
correlation of the errors of the systems to the non-correctible errors made by the users. Other
work deals with the creation of corpora of error translations and translation error correction,
such as Potet et al. (2012) and Wisniewski et al. (2013b). Other research focuses on the
creation of tools for either quality estimation such as Quest ++ (Specia et al., 2015), or quality
assessment with Kantan's MT LQR and PET (Aziz et al., 2012). Other research focuses on the
creation of tools for either quality estimation such as Quest ++ (Specia et al., 2015), or quality
assessment with Kantan's MT LQR and PET (Aziz et al., 2012). Other research focuses on the
creation of tools for either quality estimation such as Quest ++ (Specia et al., 2015), or quality
assessment with Kantan's MT LQR and PET (Aziz et al., 2012).

It is in this last vein that we position ourselves. We want to analyze translation errors on
different corpora, in collaboration with people from various backgrounds. Since the task of
analyzing errors is already tedious, we have created an online platform that can be accessed at
http://lig-accole.imag.fr. In the rest of this article, we will present our platform ACCOLÉ,
before presenting the annotated corpora obtained so far.

2 ACCOLÉ, a platform to annotate translation errors

2.1 Purpose
The initial goal that guided the development of ACCOLÉ is the manual annotation of
translation errors according to linguistic criteria. The underlying idea is to be able to provide a
user with help in choosing a MT system to use, depending on the context (language and
computer skills of the user, knowledge of the domain of the source document to be translated
and the task for which s/he needs to translate the source document.) To do this, ACCOLÉ
must be able to detect which linguistic phenomena are not handled correctly by the MT
system studied. We offer on the same platform a range of services to meet the needs of
analysis of translation errors. Thus, the main features of the platform ACCOLÉ are the
simplified management of corpora, types of errors, annotators, etc. ; annotation of errors;
collaboration and/or supervision during the annotation; the search for error models (type of
errors at first, later morphosyntactic patterns) in the annotations and finally corpora creation.

Since the task of analyzing errors is already tedious, it is important that the people carrying
it out have a simple access to the tool as well as the corpus they wish to analyze. This is why
the platform is available online on a browser and does not require any specific set up.

The task consists in labelling translation errors by locating manually their occurrences in the
target as well as in the source sentence.

2.2 Annotation Project Management
An annotation task is described on the platform thanks to the notion of project. A project is a
couple consisting of a corpus and an error typology. Thus, as shown on Figure 1, a corpus can
be associated with several error typologies in the form of several annotation projects. The
corpus will only be loaded once on the platform.

The project leader will also provide the list of annotators and supervisors. The error
typologies are handled by the platform managers.

This notion of project allows not to duplicate corpora.
As presented in table 1, two typologies of errors i.e. Vilar et al. (2006) and DQF-MQM (Lommel and Melby, 2018) are available on ACCOLÉ. Fifteen French-English corpora, journalistic news, technical documents, patents, extracts from BTEC (Basic Travel Expression Corpus) climate documents are part of the nineteen existing projects.

<table>
<thead>
<tr>
<th>Projects</th>
<th>Error Typology</th>
<th>DQF-MQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cit1</td>
<td>Citi1-Vilar/Demonstration</td>
<td></td>
</tr>
<tr>
<td>Cit2</td>
<td>Citi2-Vilar</td>
<td></td>
</tr>
<tr>
<td>News 156 Moses</td>
<td>News 156 Moses - Vilar</td>
<td></td>
</tr>
<tr>
<td>News 156 Google</td>
<td>News 156 Google - Vilar</td>
<td></td>
</tr>
<tr>
<td>News 2000</td>
<td>News 2000 - Vilar</td>
<td></td>
</tr>
<tr>
<td>BTEC1-LIGMoses</td>
<td>BTEC1-LIGMoses - Vilar</td>
<td></td>
</tr>
<tr>
<td>BTEC1-NMTLIG</td>
<td>BTEC1-NMTLIG - Vilar</td>
<td></td>
</tr>
<tr>
<td>WIPOTranslateN</td>
<td>Hawaii - Brad Pitt</td>
<td></td>
</tr>
<tr>
<td>WIPOTranslateS</td>
<td>Madagascar - George Clooney</td>
<td></td>
</tr>
<tr>
<td>CANADA_cTranslation</td>
<td>Nouvelle Calédonie - Jude Law</td>
<td></td>
</tr>
<tr>
<td>CANADA_LegacyMT@EC</td>
<td>New Zealand - Robert</td>
<td></td>
</tr>
<tr>
<td>Doc1_Fujitsu_180118 (monolingue)</td>
<td>Altica test 1</td>
<td></td>
</tr>
<tr>
<td>Doc2_Fujitsu_20180109 (monolingue)</td>
<td>Altica test 2</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Project composition

2.3 Error Typology Management

As well as permitting to upload new corpora, the platform offers the possibility to enter other error typologies.

As illustrated in figure 2, an error type will consist of a name, a category (optional), a subcategory (optional) and a code (keyboard shortcut that can be used when annotating).
ACCOLÉ offers Vilar and DQF-MQM typologies. We are planning to add MeLLange.

2.4 Error Annotation

Once in a project, one can visualize the source and target sentences and identify the error manually by clicking on the source occurrence that has not been correctly translated, and on the corresponding target error occurrence. Figure 3 shows the interface proposed to the annotator.

Figure 2. Error Typology Creation.

Figure 3. Error Annotation Process.
The annotation is done in two steps.

The first step is to select, with the mouse, words in the source sentence, and their equivalent in the target sentence, presenting a translation error. Single or Multiple selection can be done. The selection of non-adjacent words is possible. One can also select a word and its position in the translation hypothesis.

The second step is to choose the type of errors either with the help of the mouse, or with the aid of keyboard shortcuts, to be associated to the pair of previously selected source/target words. It is important to associate the type of errors with both the source and the target in order to identify error models that will ultimately propagate annotations to other corpora.

### 2.5 A collaborative platform

To answer the inter-annotator agreement issue (Popović, 2018), ACCOLÉ offers two mechanisms to help the annotator in her/his task: a supervisory mechanism allowing a manager to control the progress of the task, this mechanism encourages especially the communication between supervisor and annotator by the possibility of creating threads for a specific source/target phrase pair (asking for clarification on an error type, pointing out an annotation error, etc.); and a collaborative mechanism allowing annotators to communicate about a precise source/target phrase pair among themselves. This latest mechanism is an option to be activated in the project while creating it.

### 2.6 SSTC representation of errors

The platform uses a data representation based on the Structured String-Tree Correspondences (SSTC) (Boitet and Zaharin, 1988). An error consists of a label and a set of SNODEs (range representing the substring in the corresponding source or target phrase). For example, in Figure 4, the error concerning "all" and "any" is described by its label Poor lexical choice (cat. Object word, sub-cat. Sense), by its position in the source sentence (SNODE [49-56] - substring between the 49th and 56th characters) and the target phrase (SNODE [46-48]). The advantage of using SNODEs in this way is to avoid the use of a syntactic structure to describe the error.

Figure 4. Annotation Representation Example.

SNODEs also permit to spot the position of a missing word in the target, as we can see in Figure 1 with “un des” not translated, thus represented by SNODE[16-18] in the target. They reversely permit to associate a position in the source to extra words appearing in the target.

The other advantage is that it permits to add a posteriori morphosyntactic analysis on the source and target sentences. One or more analyzes (Stanford Tagger, Xerox Incremental Parser, etc.) can be attached using SNODE to phrases. At the end of the annotation task, its analyzes will be used to search for models of errors (morphosyntactic patterns for example). The idea is to use errors as a pivotal representation in the search mechanism.
2.7 Looking for Errors

To study the behaviour of MT systems, morphosyntactic analyses or error types (as illustrated in figure 5) are used to search for error models (morphosyntactic patterns for example) within a project. The errors are therefore used as a pivotal representation in the search mechanism. It will help looking for any consistency in the errors of MT systems.

![Figure 5. Example of an error type search.](image)

2.8 Corpora

At the time of writing this article, ACCOLÉ proposes 2 typologies of errors (those of Vilar et al (2006) and DQF-MQM (Lommel and Melby, 2018) as well as 15 French-English corpora (ranging from journalistic news, technical documents, patents, French monolingual texts, excerpts from the Basic Travel Expression Corpus (BTEC), up to climate documents), which led to the creation of 19 projects. These correspond to 6,817 sentences, 134,273 source words, 114,511 target words, for 23,525 annotations made by native English annotators, as shown in table 3.

<table>
<thead>
<tr>
<th>Projects</th>
<th>Corpora</th>
<th>Error Typologies</th>
<th>Phrase #</th>
<th>Word # (source/target)</th>
<th>Annotat ion #</th>
</tr>
</thead>
<tbody>
<tr>
<td>citi1 - Vilar</td>
<td>citi1</td>
<td>Vilar</td>
<td>516</td>
<td>13218 / 11857</td>
<td>3265</td>
</tr>
<tr>
<td>citi2 - Vilar</td>
<td>citi2</td>
<td>Vilar</td>
<td>1382</td>
<td>24402 / 21320</td>
<td>5569</td>
</tr>
<tr>
<td>News 156 LIG-Moses - Vilar</td>
<td>News 156 Moses</td>
<td>Vilar</td>
<td>156</td>
<td>4648 / 3831</td>
<td>982</td>
</tr>
<tr>
<td>News 156 Google - Vilar</td>
<td>News 156 Google</td>
<td>Vilar</td>
<td>156</td>
<td>4648 / 3715</td>
<td>616</td>
</tr>
<tr>
<td>BTEC1 LIG-Moses - DQF-MQM</td>
<td>BTEC1-LIGMoses</td>
<td>DQF-MQM</td>
<td>469</td>
<td>3445 / 3048</td>
<td>0</td>
</tr>
<tr>
<td>BTEC1 NMT LIG - DQF-MQM</td>
<td>BTEC1-NMTLIG</td>
<td>DQF-MQM</td>
<td>469</td>
<td>3445 / 3004</td>
<td>0</td>
</tr>
<tr>
<td>BTEC1 LIG-Moses - Vilar</td>
<td>BTEC1-LIGMoses</td>
<td>Vilar</td>
<td>469</td>
<td>3445 / 3048</td>
<td>807</td>
</tr>
<tr>
<td>BTEC1 NMT LIG - Vilar</td>
<td>BTEC1-NMTLIG</td>
<td>Vilar</td>
<td>469</td>
<td>3445 / 3004</td>
<td>801</td>
</tr>
</tbody>
</table>
These corpora are structured according to SNODEs and are available on request in XML or JSON format.

3 Conclusion and further work

ACCOLÉ allows the annotation of translation errors according to built-in error typologies or uploaded ones, on several annotated corpora of different texts, translated by different Statistical or Neural MT systems. ACCOLÉ offers 15 corpora, 19 projects consisting in 248,784 words and 23,525 annotations.

Some annotated corpora available have already been used for a linguistic comparison of the translation quality of different MT systems (Esperança-Rodier and Becker, 2018).

A study is planned to use ACCOLÉ for the bilingual annotation of MWEs. The error model (morphosyntactic patterns) search in the annotations, essential asset of ACCOLÉ, will allow us in a future version, to use the models of errors in order to propagate, thanks to neural network, the annotations of errors in a semi-automatic way on new corpora. This will allow to build more rapidly annotated corpora and consequently assessing almost in real-time new MT systems. We will then be able to compare the behaviour of MT systems on different domains, thus providing to the community error-annotated bilingual parallel corpora.

We also want to improve it by also allowing the association of trees of dependencies.

Further work evaluating inter-annotator agreement is still being conducted at the time of writing this paper.

Acknowledgements

The work reported above has been granted by LIG/Emergence funding in 2017 and 2019.

References


Wisniewski, Guillaume, Anil Kumar Singh, Natalia Sega and François Yvon. 2013a. Un corpus d’erreurs de traduction. TALN